Edited by
Friedlieb Pfannkuch
and Laura Suter-Dick

Predictive Toxicology

Methods and Principles in Medicinal Chemistry

Edited by R. Mannhold, H. Kubinyi, G. Folkers

Editorial Board

H. Buschmann, H. Timmerman, H. van de Waterbeemd, T. Wieland

Previous Volumes of this Series:

Kirchmair, Johannes (Ed.)

Drug Metabolism Prediction

2014

ISBN: 978-3-527-33566-4

Vol. 63

Vela, José Miguel / Maldonado, Rafael / Hamon, Michel (Eds.)

In vivo Models for Drug Discovery

2014

ISBN: 978-3-527-33328-8

Vol. 62

Liras, Spiros / Bell, Andrew S. (Eds.)

Phosphodiesterases and Their Inhibitors

2014

ISBN: 978-3-527-33219-9

Vol. 61

Hanessian, Stephen (Ed.)

Natural Products in Medicinal Chemistry

2014

ISBN: 978-3-527-33218-2

Vol. 60

Lackey, Karen / Roth, Bruce (Eds.)

Medicinal Chemistry Approaches to Personalized Medicine

2013

ISBN: 978-3-527-33394-3

Vol. 59

Brown, Nathan (Ed.)

Scaffold Hopping in Medicinal Chemistry

2013

ISBN: 978-3-527-33364-6

Vol. 58

Hoffmann, Rémy / Gohier, Arnaud / Pospisil, Pavel (Eds.)

Data Mining in Drug Discovery

2013

ISBN: 978-3-527-32984-7

Vol. 57

Dömling, Alexander (Ed.)

Protein-Protein Interactions in Drug Discovery

2013

ISBN: 978-3-527-33107-9

Vol. 56

Kalgutkar, Amit S. / Dalvie, Deepak / Obach, R. Scott / Smith, Dennis A.

Reactive Drug Metabolites

2012

ISBN: 978-3-527-33085-0

Vol. 55

Brown, Nathan (Ed.)

Bioisosteres in Medicinal Chemistry

2012

ISBN: 978-3-527-33015-7

Vol. 54

Edited by Friedlieb Pfannkuch and Laura Suter-Dick

Predictive Toxicology

From Vision to Reality

WILEY-VCH
Verlag GmbH & Co. KGaA

Series Editors

Prof. Dr. Raimund Mannhold
Rosenweg 7
40489 Düsseldorf
Germany
mannhold@uni-duesseldorf.de

Prof. Dr. Hugo Kubinyi
Donnersbergstrasse 9
67256 Weisenheim am Sand
Germany
kubinyi@t-online.de

Prof. Dr. Gerd Folkers
Collegium Helveticum
STW/ETH Zurich
8092 Zurich
Switzerland
folkers@collegium.ethz.ch

Volume Editors

Prof. Dr. Friedlieb Pfannkuch
Steingrubenweg 160
4125 Riehen
Switzerland

Prof. Dr. Laura Suter-Dick
Fachhochschule Nordwestschweiz
Gründenstrasse 40
4132 Muttenz
Switzerland

Cover has been created with pictures from Fotolia, Getty (Photodisc), Scorpius, Wikipedia, Schulz.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.

Print ISBN: 978-3-527-33608-1
ePDF ISBN: 978-3-527-67419-0
ePub ISBN: 978-3-527-67420-6
Mobi ISBN: 978-3-527-67421-3
oBook ISBN: 978-3-527-67418-3

Cover Design Grafik-Design Schulz, Fußgönheim, Germany
Typesetting Thomson Digital, Noida, India
Printing and Binding Markono Print Media Pte Ltd, Singapore

Printed on acid-free paper

Contents

List of Contributors

Arnd Brandenburg
Genedata AG
Business Unit – Expressionist
Margarethenstrasse 38
4053 Basel
Switzerland

Robert E. Burrier
Stemina Biomarker Discovery, Inc.
504 South Rosa Road
Madison, WI 53719
USA

Antonello Caruso
Roche Pharma Research and Early Development (pRED)
Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Elizabeth L.R. Donley
Stemina Biomarker Discovery, Inc.
504 South Rosa Road
Madison, WI 53719
USA

Laura A. Egnash
Stemina Biomarker Discovery, Inc.
Client Services and Operations
504 South Rosa Road
Madison, WI 53719
USA

Heidrun Ellinger-Ziegelbauer
Bayer Pharma AG
Investigational Toxicology
Aprather Weg 18a
42096 Wuppertal
Germany

Brigitte Faquet
L'Oréal Research & Innovation
Predictive Model and Method Development Department
1 Avenue Eugène Schueller
93601 Aulnay-sous-Bois
France

Nicole Flamand
L'Oréal Research & Innovation
Predictive Model and Method Development Department
1 Avenue Eugène Schueller
93601 Aulnay-sous-Bois
France

Timothy W. Gant
Public Health England
Centre for Radiation, Chemical and Environmental Hazards
Harwell Science and Innovation Campus
Didcot
Oxfordshire OX11 0RQ
UK

Hans Gmuender
Genedata AG
Business Unit – Expressionist
Margarethenstrasse 38
4053 Basel
Switzerland

Hans Peter Grimm
Roche Pharma Research and Early Development (pRED)
Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Tobias Heckel
F. Hoffmann-La Roche Ltd
Pharma Early Research and Development (pRED) – Innovation Center Basel
Pharmaceutical Sciences (PS)
Grenzacherstrasse 124
4070 Basel
Switzerland

Philip Hewitt
Merck KGaA
Non-Clinical Safety
Frankfurterstrasse 250
D-64283 Darmstadt
Germany

Lisa Hoelting
University of Konstanz
Department of Biology, Research School Chemical Biology
Universitätsstraße 10
78457 Konstanz
Germany

Paul Jennings
Innsbruck Medical University
Department of Physiology and Medical Physics
Division of Physiology
Fritz-Pregl-Straße 3
6020 Innsbruck
Austria

Esther Johann
Merck KGaA
Non-Clinical Safety
Frankfurterstrasse 250
D-64283 Darmstadt
Germany

Harald Kropshofer
F. Hoffmann-La Roche Ltd
Pharmaceutical Division
Grenzacherstrasse 124
4070 Basel
Switzerland

Doris Lagache
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Thierry Lavé
Roche Pharma Research and Early Development (pRED)
Project leaders/Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Marcel Leist
University of Konstanz
Department of Biology
Universitätsstraße 10
78457 Konstanz
Germany

Martin O. Leonard
Public Health England
Centre for Radiation, Chemical and Environmental Hazards
Harwell Campus
Harwell Oxford Science Park
Chilton, Didcot
Oxfordshire OX11 0RQ
UK

Angela Mally
University of Würzburg
Department of Toxicology
Versbacher Strasse 9
97078 Würzburg
Germany

Emma L. Marczylo
Public Health England
Centre for Radiation, Chemical and Environmental Hazards
Harwell Campus
Harwell Oxford Science Park
Chilton, Didcot
Oxfordshire OX11 0RQ
UK

Astrid Mayoux
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Christophe Meille
Roche Pharma Research and Early Development (pRED)
Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Natalie Mesens
Janssen Pharmaceutical Companies of Johnson & Johnson
Discovery Support and Investigative Toxicology
Preclinical Development & Safety
Turnhoutseweg 30
2340 Beerse
Belgium

Lauren Nakab
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Fabrice Nesslany
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Gladys Ouédraogo
L'Oréal Research & Innovation
Predictive Model and Method Development Department
1 Avenue Eugène Schueller
93601 Aulnay-sous-Bois
France

Jessica A. Palmer
Stemina Biomarker Discovery, Inc.
Cell Biology Department
504 South Rosa Road
Madison, WI 53719
USA

Neil Parrott
Roche Pharma Research and Early Development (pRED)
Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Friedlieb Pfannkuch
Toxconsult
Steingrubenweg 160
4125 Riehen
Switzerland

Beatriz Silva Lima
Universidaded de Lisboa
Faculdade de Farmácia
iMED.ULisboa
Avenida Professor Gama Pinto
1649-003 Lisbon
Portugal

Sophie Simar
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Thomas Steger-Hartmann
Bayer HealthCare
Investigational Toxicology
Müllerstrasse 178
13353 Berlin
Germany

Luc Stoppini
University of Applied Sciences Western Switzerland
Tissue Engineering Laboratory
Haute école du paysage, d'ingénierie et d'architecture de Genève – hepia
Rue de la Prairie 4
1202 Geneva
Switzerland

Laura Suter-Dick
University of Applied Sciences North-Western Switzerland
High School of Life Sciences – Molecular Toxicology
Gründenstrasse 40
4132 Muttenz
Switzerland

Smail Talahari
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Eric Vercauteren
Institut Pasteur de Lille
Laboratoire de toxicologie
1 rue du Professeur Calmette
59019 Lille Cedex
France

Antje-Christine Walz
Roche Pharma Research and Early Development (pRED)
Modeling & Simulation, Pharmaceutical Sciences
Roche Innovation Center Basel
F. Hoffmann-La Roche Ltd
Grenzacherstrasse 124
4070 Basel
Switzerland

Anja Wilmes
Innsbruck Medical University
Department of Physiology and Medical Physics
Division of Physiology
Fritz-Pregl-Straße 3
6020 Innsbruck
Austria

Timo Wittenberger
Genedata AG
Business Unit – Expressionist
Margarethenstrasse 38
4053 Basel
Switzerland

Preface

With great pleasure we announce volume 64 of our book series "Methods and Principles in Medicinal Chemistry." The volume editors Friedlieb Pfannkuch and Laura Suter-Dick present an excellent book dedicated to predictive toxicology, a highly important research area with prime impact on the quality of compounds from drug discovery and development projects. Therapeutic use of any new compound is in demand of a thorough identification and profiling of its safety. Protection of human safety is a primary objective of toxicology research and risk assessment.

Toxicology is the study of the adverse effects of drugs and other chemicals on living systems and the means to prevent or at least minimize such effects. Toxicology is a multifaceted field, overlapping with biochemistry, histology, pharmacology, pathology, and several others. Subdisciplines of toxicology include clinical, regulatory, forensic, and occupational toxicology as well as risk assessment.

Poor pharmacokinetics, side effects, and compound toxicity are frequent causes of late-stage failures in drug development. A safe *in silico* identification of adverse effects triggered by drugs and chemicals would be highly desirable as it not only bears economic potential but also spawns a variety of ecological benefits.

The drug development industry has undertaken significant efforts to identify toxic events at the earliest opportunity during the development process, moving from a predominantly observational science at the level of disease-specific models to a more predictive model focused on target-specific mechanism-based biological observations. The growth in such Early Safety Assessment initiatives has driven the need for more reliable, cost-effective high-throughput *in vitro* toxicity assays capable of predicting toxic liabilities prior to investment in more costly preclinical and clinical trials.

In silico toxicology studies can help to focus *in vitro* and *in vivo* experiments to make the latter highly efficient. In some cases, *in silico* studies might even replace particularly expensive, lengthy, uninformative, or offensive *in vitro* or *in vivo* experiments. Moreover, by virtue of being computer-based and, hence, inexpensively replicable, *in silico* toxicology can vastly expand the applicability and availability of toxicological analysis [1–6].

The ultimate goal for predictive toxicology would be the ability to go from visualizing the chemical compound structure to predicting its safety profile. The major

challenge is to translate the tremendous scientific progress in this field into practical use or general acceptance. Scientists are using biological data very effectively – whether it is gene expression data or even data from proteomic or other profiling techniques – to gain a sense of whether a drug is having off-target effects or otherwise adversely impacting the system. As technologies become more mechanism-based and as more data accrue, it should enable predictions with better accuracy and decrease occurrences of false negatives and false positives.

Chapters of this comprehensive volume consider all topic areas relevant in the field of predictive toxicology, such as *in silico* approaches, data management, and bioinformatics (Chapters 2–4), omics technologies and biomarker development (Chapters 5–10), advanced *in vitro* systems (Chapter 11), models for cosmetic products (Chapter 12), use of stem cells with focus on neurotoxicology and teratology (Chapters 13 and 14), immunogenicity of protein therapeutics (Chapter 15), and finally aspects on acceptance by Drug Regulatory Authorities (Chapter 16).

The series editors are grateful to Friedlieb Pfannkuch and Laura Suter-Dick for organizing this volume and collaborating with excellent authors. Last but not least, we thank Frank Weinreich and Heike Nöthe from Wiley-VCH for their valuable contributions to this project and to the entire book series.

Düsseldorf
Weisenheim am Sand
Zürich
October 2014

Raimund Mannhold
Hugo Kubinyi
Gerd Folkers

References

1 Davis, M., Boekelheide, K., Boverhof, D.R., Eichenbaum, G., Hartung, T., Holsapple, M.P., Jones, T.W., Richard, A.M., and Watkins, P.B. (2013) The new revolution in toxicology: the good, the bad, and the ugly. *Ann. N.Y. Acad. Sci.*, **1278**, 11–24.

2 Ekins, S., Mestres, J., and Testa, B. (2007) *In silico* pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol.*, **152**, 21–37.

3 Sturla, S.J., Boobis, A.R., FitzGerald, R.E., Hoeng, J., Kavlock, R.J., Schirmer, K., Whelan, M., Wilks, M.F., and Peitsch, M.C. (2014) Systems toxicology: from basic research to risk assessment. *Chem. Res. Toxicol.*, **27**, 314–329.

4 Modi, S., Hughes, M., Garrow, A., and White, A. (2012) The value of *in silico* chemistry in safety assessment of chemicals in the consumer goods and pharmaceutical industries. *Drug Discov. Today*, **17**, 135–142.

5 Niklas, N., Diaz Ochoa, J.G., Bucher, J., and Mauch, K. (2013) Quantitative evaluation and prediction of drug effects and toxicological risk using mechanistic multiscale models. *Mol. Inform.*, **32**, 14–23.

6 Ortwine, D.F. and Aliagas, I. (2013) Physicochemical and DMPK *in silico* models: facilitating their use by medicinal chemists. *Mol. Pharm.*, **10**, 1153–1161.

A Personal Foreword

Knowingly or unknowingly, toxicology affects most parts of our society. There is clear public interest in only accepting products with a well-characterized safety profile in the market. There is also a requirement for several industries such as pharmaceutical, chemical, and cosmetic industries to perform a battery of *in vitro* and animal studies in order to avoid harm to the general public, volunteers in clinical trials, patients, workers in production plants, and the environment.

Toxicology is a multidisciplinary science for evaluation of risk/benefit ratio and takes its methods from other sciences such as chemistry and pharmacological chemistry, pharmacology, pathology, biochemistry, clinical medicine, and forensic medicine.

The spectrum of toxicity assays comprises computational (*in silico*) methods as well as the testing of chemicals with *in vitro* methods and in selected laboratory animal species to describe the dose–effect relationship over a broad range of doses in order to detect secondary (harmful/unwanted) pharmacological effects and adverse (toxic) effects.

The final goal must be the extrapolation and prediction of adverse effects to humans. The challenge is to identify a safe dose in humans and setting exposure limits (ceiling), if required. In this context, potential target organs of toxicity and reversibility of potential side effects should be identified and meaningful parameters for (clinical) monitoring should be chosen. Finally, the discipline should contribute to the elucidation of mechanisms of toxic/adverse effects.

Industry's activities are driven by national, regional, and global regulatory requirements [1–4], strategic and commercial aspects, and scientific and technological state of the art. These aspects are the main driving forces for advances in toxicology. In addition, there is an increasing public pressure to refine, reduce, and replace animal testing ("3Rs" [5]) for ethical reasons.

The scientific and regulatory environment is changing rapidly. The introduction of new technologies (e.g., for testing of biologics or new approaches to improve carcinogenicity testing) and the trend toward perfectionism (e.g., including as many investigation parameters as possible) have caused the extension of the existing study programs and a dramatic increase in the investment of human and financial resources.

However, we have relied for decades on the use of animal studies (*in vivo* toxicology) with generally unsatisfactory predictive performance (acknowledged by many and summarized by Olson *et al.* [6]). In particular for the pharmaceutical industry, several products have caused serious adverse reactions despite having been through a battery of mandatory toxicity tests. The consequences of this suboptimal predictive performance are often disastrous for the patients and for the pharmaceutical industry.

Thus, for the past few decades, predictive approaches other than studies in animals have been considered and employed with varying degrees of success. Among the more commonly used approaches are *in silico* tools, *in vitro* assays with primary cells and cell lines combined with specific endpoints, omics technologies, and the use of stem cell-derived cells.

The diversity of technologies and scientific knowledge that flows into the advanced approaches used currently to predict toxicity require a new type of biologist, different from the one traditionally recruited for performing toxicology testing. They must have an in-depth knowledge of the applied technological advancements, biological networks, and adverse outcome pathways, and, most importantly, a thorough understanding of the contextual relevance of the biological findings in relation to toxicology and pathology, and ultimately to the effect on the human population.

Any future activity, however, must focus on the improvement of the predictivity of toxicology/safety testing, since identification of potential safety issues upstream in the drug discovery process is a major bottleneck in drug development. New technologies may play a central role in this respect.

The final goal must be to combine the results from new technologies and classical toxicology methodology in a scientifically sound way in order to gain acceptance by Regulatory Authorities and we strongly hope that this book is contributing to this challenge.

The aim of this book is to provide a comprehensive overview of the latest scientific developments in the field of "predictive" toxicology and their applications in safety assessment. The topics have been tackled by selected expert scientists, who are familiar with the theoretical scientific background as well as with the practical application of methods and technologies. To ensure scientific excellence related with practical application of the contributions, we have invited scientists active both in the academic and in the industrial toxicology research.

Finally, we want to acknowledge the pleasant collaboration with Dr. Heike Noethe and Dr. Frank Weinreich from Wiley-VCH for their constant support during all steps of editing this book.

Basel
October 2014

Friedlieb Pfannkuch and Laura Suter-Dick

References

1 European Community (EC) (2014) The Rules Governing Medicinal Products in the European Community. Available at http://ec.europa.eu/health/documents/eudralex/index_en.htm (last assessed February 2014).

2 FDA (2013) Guidelines for the Format and Content of the Nonclinical Pharmacology/Toxicology Section of an Application. US Department of Health and Human Services, Public Health Service, Food and Drug Administration, Washington, DC. Available at http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/NewDrugApplicationNDA/ (last assessed February 2014).

3 Japanese Ministry of Health and Welfare (1993) Guidelines for Toxicity Studies of Drugs. Notification No. 88 of the Pharmaceutical Affairs Bureau of the Ministry of Health and Welfare. Available at http://www.jpma.or.jp/english/parj/pdf/2012.pdf (last assessed February 2014).

4 European Union (2014) ICH Guidelines. Available at www.ich.org (last assessed February 2014).

5 Russell, W.M.S. and Burch, R.L. (1959) BT The Principles of Humane Experimental Technique, Methuen & Co. Ltd, London (reissued 1992). Available at http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc (last assessed February 2014).

6 Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Dorato, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.*, **32** (1), 56–67.

1
Introduction to Predictive Toxicology Tools and Methods

Laura Suter-Dick and Friedlieb Pfannkuch

1.1
Computational Tools and Bioinformatics

1.1.1
In Silico Prediction Tools

Computational tools are used in many life sciences research areas, including toxicity prediction. They take advantage of complex mathematical models to predict the effects caused by a given compound on an organism. Due to the complexity of the possible interactions between a treatment and a patient and the diversity of possible outcomes, models are applied to well-defined and specific fields, such as DNA damaging potential, estimation of the necessary dose to elicit an effect in a patient, or identification of relevant gene expression changes.

In silico tools make use of information regarding chemical structures and the immense data legacy that allows inferring interactions between chemical structures, physicochemical properties, and biological processes. These methods are farthest away from traditional animal studies, since they rely on existing databases rather than on generating experimental animal data.

Due to the complexity of this task, there are a fairly small number of endpoints that can be predicted with commonly employed *in silico* tools such as DEREK, VITIC, and M-Case with acceptable accuracy. In order to improve the current models and to expand to additional prediction algorithms, further validation and extension of the underlying databases is ongoing.

Similarly, modeling and simulation (M&S) can generate mathematical models able to simulate and therefore predict how a compound will behave in humans before clinical data become available. In the field of nonclinical safety, complex models allow for a prediction of the effect of an organism on a compound (pharmacokinetic models) as well as, to some extent, pharmacodynamic extrapolations, based on data generated in animal models as well as in *in vitro* human systems.

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

1.1.2
Bioinformatics

In addition to the *in silico* and modeling tools described above, the dramatically increasing amount of toxicologically relevant data needs to be appropriately monitored and collected. All "new" technologies produce very high volumes of data and thus having and using bioinformatics tools that can collect data from diverse sources and mine them to detect relevant patterns of change is vital. For this purpose, large databases are necessary, along with bioinformatics tools that can deal with diverse data types, multivariate analysis, and supervised and unsupervised discrimination algorithms. These tools take advantage of advanced statistics, combined with the large data sets stored in the databases generated using technologies such as omics or high-content imaging.

1.2
Omics Technologies

The omics technologies arose with the advent of advanced molecular biology techniques able to determine changes in the whole transcriptome, proteome, or metabolome. These powerful techniques were considered the ultimate holistic approach to tackle many biological questions, among them toxicological assessment. Several companies have invested in these areas of toxicological research.

1.2.1
Toxicogenomics (Transcriptomics)

Toxicogenomics is the more widespread of the omics technologies. Predictive approaches are based on databases with compounds (toxic/nontoxic) generated by (pharmaceutical) companies as well as by commercial vendors in the 1990s. All share the same focus of investigation: target organ toxicity to the liver and the kidney.

In addition, gene expression data are often the basis for mechanistic understanding of biological processes in several fields, including toxicology, pharmacology, and disease phenotype. Thus, transcriptomic data can be used as a merely predictive tool, as a mechanistic tool, or as a combination of both.

Subsequent to global gene expression analysis, assays can then be developed for a relatively small subset of genes relevant to specific toxicities and toxic mechanisms. Such assays can be used for screening and problem solving in toxicology studies and may also play a role in efficacy studies in preclinical and clinical research.

In addition, the modulation of gene expression through toxicants has also been used as a method for the discovery of putative biomarkers. In particular in

the kidney, gene expression changes detected after renal injury led to protein assays in urine that could be measured noninvasively.

In the last few years, the ease with which full genome sequencing and mRNA sequencing can be performed has led to other ways to analyze samples and tissues. Sequence information from different species is also more readily available and allows for better interpretation of genomic and proteomic data, in particular when extrapolating to humans. The usefulness of this approach to predict the drug effects in humans and the latest developments in the field, the next-generation sequencing (NGS), will be discussed in this book in more detail.

Also, the role of DNA methylation, and of a variety of small noncoding RNA molecules, such as miRNAs, is gaining importance in toxicological assessment and biomarker discovery.

1.2.2 Proteomics

Proteomics is probably the oldest of the omics technologies and arguably the most relevant from a biological point of view, since protein expression and protein posttranslational modifications are the executors of cellular processes. From a knowledge point of view, there are large, publicly available databases with protein sequences that can be used for the identification of proteins.

However, proteomics is also probably the most technologically challenging of the three omics, mainly due to the large diversity in proteins, in particular in terms of abundance and physicochemical properties. This poses a massive challenge since it requires a technology with a dynamic range of several orders of magnitude as well as separation methods able to deal with extreme differences in lipophilicity.

Protein expression changes (or modifications such as phosphorylation) that can be detected are valuable pieces of information and can be used to understand biological pathways and to discover new biomarkers.

1.2.3 Metabolomics

Similarly to toxicogenomic databases, metabolomic databases were generated by pharmaceutical companies, academia, and commercial providers. Metabolomic data were mainly generated using ^{1}H NMR and chromatography coupled with mass spectroscopy methods. The most commonly used fluids are urine and plasma, although tissue or cellular extracts can also be analyzed.

The main advantage of metabolomics is that the sampling of body fluids can be performed noninvasively. Thus, there were high expectations of using metabolomics for the discovery of new translational biomarkers. However, delivery has been slower than anticipated and much of the research to date has been descriptive rather than predictive [1].

1.3
Data Interpretation and Knowledge Management

An increasing number of molecular biology technologies related to toxicology research are available. In addition to genes, proteins, and metabolites, we have now the means to analyze many other factors that regulate and/or influence the expression of genes and proteins, as well as the secretion of metabolites. For example, it has been recognized that we should pay closer attention to DNA methylation status and miRNA expression, factors that profoundly regulate gene and protein expression, respectively.

Regarding the interpretation of the data in the context of safety assessment, there is still a lack of understanding of changes that occur during normal adaptive variations in physiology as opposed to changes due to alterations in pharmacology, so toxicity-related changes can be difficult to untangle.

The concept of "adverse outcome pathway" currently indicates the interest in identifying changes that will lead to a clinically relevant effect. To this end, it is also becoming apparent that the greatest benefit can be obtained by integrating these newer technologies with information from conventional toxicology and pathology.

1.4
Biomarker Development

The integration of highly sensitive molecular biology technologies with the well-established pathology assessment also provides a means to identify putative novel biomarkers. These markers not only may indicate toxicity but can also be used as markers for specific disease conditions.

Ideal biomarkers would allow monitoring onset, progression, and reversibility of adverse events and be translational, enabling their use in both preclinical and clinical settings [2–4].

For safety assessment, it is a major endeavor of toxicology research to identify sensitive and specific biomarkers, ideally translational across species and prodromal, for example, able to predict toxicities that may arise after prolonged exposure. As indicated above and although the omics technologies can identify many candidate markers, the amount of work and time required to validate these putative biomarkers is extremely challenging.

1.5
Advanced *In Vitro* Systems and Stem Cell Research

1.5.1
Advanced *In Vitro* Testing

In vitro toxicology is not a new concept and the advantages of cell cultures are manifold, including the possibility of screening at low cost and the reduction in

animal experimentation, supporting the 3Rs concept (refine, reduce, and replace animal experiments).

However, in the same way that the determination of an LD_{50} in animals is not really a sensitive and meaningful endpoint for side effects in humans, simple cytotoxicity assays will not be very informative to the outcome in a whole organism. This widely accepted fact has led to the development of a plethora of *in vitro* systems that can be applied in toxicology. These systems are designed to provide mechanistic answers related to specific organ toxicities, rather than to determine the overall safety profile of a drug candidate or a chemical.

They span from simple bacterial cellular systems (e.g., the well-established Ames test) to novel organotypical multicellular systems that are still largely under development. These advanced *in vitro* test systems try to combine cell culture conditions that are relevant to the target organ under investigation with advanced endpoints, such as high-content imaging or molecular markers. Besides the well-known cell lines grown traditionally in monolayers, particular attention is currently being given to the tissue architecture of the cultures. The most commonly investigated systems address liver, kidney, CNS, and heart.

Two major ways of improving the relevance of the cultures are currently at work. On the one hand, major efforts are devoted to generate 2D/3D cultures with several cell types that should better mimic the physiology of the organ, including aspects such as cell–cell interactions [5]. On the other hand, a large amount of research is focused on producing and characterizing the most relevant cells. For example, some cell lines have been engineered to resemble more closely the hepatocyte.

In addition to the use of mammalian cell systems, the use of nonmammalian species such as the zebrafish has produced interesting results that can be applied to several toxicology fields. One of the advantages of using whole organisms is clearly that it provides the largest degree of complexity, since it gives access to all organs and their interplay. The small size of the zebrafish enables screening-type testing.

Using a non-mammalian animal, the extrapolation to humans becomes more difficult. However, for specific endpoints such as teratogenicity, zebrafish has proven very reliable.

A future challenge for *in vitro* toxicology for safety assessment remains the exposure levels *in vitro* and their relevance to the *in vivo* situation. Compound concentration, protein binding, and time of exposure of the cell cultures in relation to the toxicity outcome and the expected *in vivo* plasma and organ concentrations should be taken into account.

1.5.2 Stem Cell Research

Also, taking advantage of the exciting developments in stem cell research and reprogramming of somatic cells for the generation of pluripotent cells (induced

pluripotent stem cells (iPSCs)), many researchers are investigating the use of human stem cell-derived cells (embryonic stem cells (ESCs) or iPSCs) to avoid the need to extrapolate across species.

Stem cells are pluripotent; thus, they have the potential to differentiate into any cell in the organism. Until 2006, stem cells were obtained from embryos, either from animals (mainly mouse: mESCs) or from supernumerary embryos from *in vitro* fertilization (IVF) programs (human: hESCs).

The use of hESCs has technical limitations and is burdened with ethical issues. Technically, it is not possible to define or select the genotype of the hESC line and there are a limited number of sources. Ethically, it is clear that working with human embryos has brought up moral issues, leading to legal restrictions in some countries. A way to circumvent the use of hESCs was made possible in 2006, when Takahashi and Yamanaka [6] published their work on the generation of iPSCs from mouse. A year later, Thomson and coworkers published a similar method to obtain human iPSCs [7], which opened amazing opportunities for *in vitro* research as well as in regenerative medicine.

In the field of toxicology, stem cells (either ESCs or iPSCs) are being increasingly used. Some of these systems are already advanced, for example, the teratogenicity screening using mouse or human ESCs and the cardiotoxicity screening using hESC- or iPSC-derived cardiomyocytes.

Some of the major technical issues of stem cell-derived cell types concern the efficient reprogramming and the appropriate differentiation into the cell type of interest. At present, the reprogramming protocols are becoming more and more robust, but still quite some work needs to be done in the targeted differentiation, including the upscaling to produce sufficient number of cells of homogeneous quality. A major problem remains the differentiation efficiency, with sometimes only a small percentage of cells differentiating into a cell type under investigation, giving rise to hard to control cocultures. Also difficult to tackle is the adequate characterization of the differentiated cells (when do you consider them terminally differentiated?) and their stability (how long do they keep the differentiated state in culture?).

Despite these issues inherent to any new technology, great results have been published with differentiated cardiomyocytes and neuronal cells. Also, and of immense relevance for the toxicology field, hepatocyte-like cells are being made and the protocols are being optimized.

1.6 Immunogenicity

In the last decade, the biopharmaceutical drug market has expanded with a faster growth rate than that of classical "small molecule" drugs. To date, more than 20 therapeutic monoclonal antibodies (mAbs) have been released to the market for the treatment of several diseases and many more are in development.

These biopharmaceutical macromolecules pose very specific challenges to safety assessment. On the one hand, it is assumed that through their extremely high specificity, there will be no off-target activity that might lead to unspecific side effects. On the other hand, macromolecules can trigger very strong reactions in the immune system, which might lead to reduction in efficacy or to serious adverse events.

Although immune reactions in general and immunogenicity in particular are one of the major concerns when assessing the safety of macromolecules, small chemical entities can also trigger immune responses. Some of these immune reactions have been shown to lead to hepatotoxicity. The immune responses are usually triggered either by own immunogenic capacity of the small molecule (e.g., halothane) or through covalent modification of macromolecules in the host. Also, it has been postulated that an immune component underlies the so-called idiosyncratic hepatotoxicity.

1.7
Integration and Validation

1.7.1
Use of Omics for Toxicology Testing

The immense expectations of the omics technologies to solve all problems were probably naïve and have proven to be unrealistic. Some technical aspects were the first stumbling blocks. For toxicogenomics (transcriptomics), there was initially a question of its robustness with regard to the methodology and reproducibility between different platforms, but this has been largely overcome. Proteomic technologies are still not at a stage where many low-abundance proteins can be isolated and identified. The initial metabolomic experiments suffered from a lack of sensitivity and changes in the metabolic profile of fluids such as urine can be confounded by physiological variation due to changes in diet or energy metabolism during weight loss at toxic doses.

In addition to the technical aspects that have been more or less solved in the last 10–15 years, the interpretation of the data and their relevance to biological processes in a given organism remain complex. However, knowledge of technological limitations together with increased understanding of involved pathways has made omics very useful tools for mechanistic toxicology studies.

1.7.2
Integration of "New" Technologies into Risk Assessment

Despite the thrilling scientific advances, extreme care must be taken to correctly integrate the new type of data into a risk assessment process. Over the past few decades, the predictivity of toxicology has become very reliable in certain areas, for example, mutagenicity and reproduction toxicity. On the other hand, the

relevance of nonclinical results for predicting potential hazards in humans was and is fundamentally under investigation and discussion.

Introduction of any new method will depend on its progress/contribution/reliability regarding the prediction of unwanted/adverse effects in humans. Thus, the use of newer technologies must be validated against the "classical" toxicological methods or approaches. Significant staff time and effort will be needed to follow up possible "signals" and novel biomarkers.

For this purpose, senior scientists/biologists are required to help interpret the data with an open-minded, yet critical attitude, in order to assess the relevance of the data. This will certainly require additional training so that opportunities and pitfalls of the newer technologies can be recognized by all stakeholders of drug development and integrated in a realistic way.

1.7.3 Use of Human-Derived Cellular Systems

The combination of both, human-derived cells and organotypical culture systems, together with appropriate endpoints to monitor cellular processes (such as high-content imaging), may revolutionize the field of *in vitro* toxicology in the near future.

This advancement is making possible the widespread use of human material for toxicity assessment, obliterating altogether the need for extrapolation of results in different animal species to humans. Despite the fact that there are many open questions with regard to validation of such data, it can be foreseen that its relevance will strongly contribute to the safety assessment for the target of prediction, the human being.

There are few compounds with a specific toxic potential to humans only and they may thus serve as positive controls in prospective validation experiments. On the other hand, there are a number of substances that display species-specific toxicities in animals only and for which respective human data are lacking. Thus, the definition of false-positive and false-negative results may need to be revisited.

1.7.4 "General" Acceptance – Translation into Guidelines

Safety assessment of drugs and chemicals is highly regulated by national and international guidelines. The scientific officers in the regulatory authorities of the different countries or regions take the responsibility to allow only products with a well-characterized safety profile on the market/in the public.

Developing, adapting, or modifying the existing regulatory requirements/guidelines is a highly responsible task and a long-lasting process. Officers of regulatory authorities are heavily relying on solid data, and they are very interested in getting state-of-the-art results from academia and industry to base their decision on.

1.8
Research Initiative/Collaborations

The fast pace of scientific and technological advances in the field of toxicology generates amazing opportunities. However, to make the best out of this, a large amount of financial and scientific resources and time are necessary. Thus, rather than generating core information in individual research groups or companies, research can be performed in consortia, taking advantage of synergies and optimizing the use of the funding.

There has been major public funding in Europe (e.g., EU Research Framework Programs, Innovative Medicines Initiative - IMI) as well as in the United States (e.g., Critical Path Institute. The Toxicology in the 21st Century (Tox21) program, a federal collaboration involving the NIH, Environmental Protection Agency (EPA), and Food and Drug Administration (FDA).) and Japan (e.g., The Toxicogenomics Consortium, TGP). Many of the funding schemes also involve funding from industry (pharmaceutical, chemical, cosmetics, and consumer products). Thus, there is great interest of society as a whole in promoting science in support of predictive toxicology. Specific consortia are discussed in more detail in several chapters of this book.

1.9
Concluding Remarks

Over the past three decades, there was a dramatic increase in knowledge in the field of biological sciences and we are in the middle of a phase where this progress will translate into practical paradigm changes regarding toxicity testing and safety assessment of chemicals and (bio-)pharmaceuticals.

References

1 Kramer, J.A. and Kolaja, K.L. (2002) Toxicogenomics: an opportunity to optimise drug development and safety evaluation. *Expert Opin. Drug Saf.*, **1**, 275–286.

2 Tugwood, J.D., Hollins, L.E., and Cockerill, M.J. (2003) Genomics and the search for novel biomarkers of toxicity. *Biomarkers*, **8**, 79–92.

3 Bandara, S.R. and Kennedy, S. (2002) Toxicoproteomics – a new preclinical tool. *Drug Discov. Today*, **7**, 411–418.

4 Lindon, J.C. *et al.* (2003) Contemporary issues in toxicology: the role of metabonomics in toxicology and its evaluation by the COMET project. *Toxicol. Appl. Pharmacol.*, **187**, 137–146.

5 Atienzar, F.A., Eric, I., Novik, E.I., Gerets, H.H., Parekh, A., Delatour, C., Cardenas, A., MacDonald, J., Yarmush, M.L., and Dhalluin, S. (2014) Predictivity of dog co-culture model, primary human hepatocytes and HepG2 cells for the detection of hepatotoxic drugs in humans. *Toxicol. Appl. Pharmacol.*, **275** (1), 44–61.

6 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126** (4), 663–676.

7 Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318** (5858), 1917–1920.

2
In Silico Toxicology – Current Approaches and Future Perspectives to Predict Toxic Effects with Computational Tools

Thomas Steger-Hartmann

2.1
Introduction

In silico prediction tools in toxicology have significantly evolved in the last decade and a few systems for specific toxicological endpoints have found their entry into regulatory risk assessment [1]. However, rather than completely replacing both *in vitro* and *in vivo* toxicological methods [2], *in silico* tools have proven to be complementary to the standard testing approaches and an independent field of "computational toxicology" is yet not broadly established [3]. This chapter will analyze the current situation and provide a comprehensive overview of available approaches and uses of *in silico* tools in toxicology. It will also critically discuss the existing limitations and inroads to overcome these. Although the mathematical approaches are similar in the field of ecotoxicology, this overview will limit itself to the use of predictive tools in mammalian or human toxicity assessment, because deriving mechanistic conclusions and assessing exposure significantly differ between these two fields. Therefore, *in silico* tools for ecotoxicology are outside the scope of this chapter.

2.2
Prediction of Hazard

2.2.1
Definition of Hazard and Its Use

A basic question for the use and application of any toxicological method is whether it will serve for hazard identification or risk assessment. This also applies for *in silico* predictive systems. There is no unambiguous internationally accepted definition of the term "hazard" [4]; however, the broadly accepted use in toxicology refers to the intrinsic toxic properties of a compound, that is, the question whether a compound causes an adverse effect. Whereas the strength of the effect may play a role in the hazard classification (e.g., whether a compound

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

is a weak or a strong sensitizer), hazard identification is generally independent of the actual exposure and thus unrelated to concentrations or doses.[1)]

Hazard identification is sufficient for labeling compounds. For example, the Globally Harmonized System (GHS), which replaces the R-phrases, classifies the risk of acute toxicity into five hazard categories based on LD_{50} or LC_{50} values for dermal, oral, and inhalation routes of exposure. However, for the assessment of the factual risk inherent to a compound, the exposure information, that is, the applied dose or even the measured plasma levels, are indispensable. Constructing *in silico* tools for the prediction of hazard is a much simpler task than predicting risk. In the case of hazard prediction, one basically only needs to collect chemical series for which the toxicological endpoint of interest has been investigated. The chemicals and their measured effects can be sorted according to their substructures, which are known or suspected to cause the toxic effects. From the substructures, so-called "structural alerts" [5] can be defined and implemented in a computational expert system. Alternatively, the structures could be dissected into chemical descriptors that could then be statistically analyzed for correlation with the biological effect, resulting in a quantitative structure–activity relationship (QSAR) tool. Transformed into computer algorithms both approaches will result in the prediction of the hazard potential for a new compound in a binary fashion (yes or no), but will not provide substantial information whether there is a significant risk of harm to an organism if it is exposed to a specific concentration or dose. At most, such a QSAR tool will predict whether a compound will cause a weak or a strong effect, if it is constructed in a way that the chemical structures are correlated with the quantitative readout of the biological assay.

2.2.2 Prediction of Mutagenicity

The above-described binary outcome of an *in silico* prediction is sufficient in such fields of toxicology where some kind of zero tolerance policy is established. This is the case for DNA-reactive mutagens, because the toxicological theory of the so-called one-hit model assumes that only one genetic change is required to transform a normal cell into a cancer cell, and therefore a single molecule of a genotoxic (DNA-reactive) carcinogen presents a minute but finite risk of cancer [6]. The theory has experienced various modifications and the scientific debates together with profound data analysis have resulted in the introduction of the threshold of toxicological concern (TTC) even for compounds predicted to be mutagenic [7]. The TTC concept was originally developed for food contaminants using a linear extrapolation of carcinogenic potency data from rodent carcinogenicity assays. An exposure below the calculated thresholds (TTCs)

1) In a strict sense and from an experimental perspective, this only applies to mutagenicity testing, where there are no thresholds for evaluation of positive (mutagenic) effects. Many other assays for hazard identification have threshold levels that have to be passed in order to classify an effect as adverse or to attribute different toxic classes.

defines the level for any unstudied chemical that will not pose a risk of significant carcinogenicity. From these potency data sets, structural alerts were derived, which are now also applied to chemicals beyond the field of food contaminants [8].

The original theory for mutagenic effects resulting in a binary answer (mutagenic – yes or no?) still makes an *in silico* approach relatively attractive for this endpoint. Together with the good understanding of the mechanistic causes of mutagenicity and the availability of large data sets [9], this paradigm has laid the ground for the successful establishment of *in silico* tools for hazard prediction in genetic toxicology. Both expert systems (e.g., DEREK Nexus, ToxTree) and QSAR tools (e.g., MultiCase, Sarah Nexus, TopKat, Leadscope Model Applier) are used in the toxicological community for the prediction of mutagenicity. Especially in such cases where a structure represents an impurity of a manufacturing process that cannot be isolated in amounts sufficient for experimental testing, these systems meanwhile play a pivotal role. Most of these systems are constructed in a rather transparent way and can provide a certain level of validation, which led to the acceptance of such tools in the regulatory assessment of genotoxic impurities [10,11]. Prediction of DNA-reactive impurities currently remains the only situation where the outcome of an *in silico* tool alone determines regulatory decision making, that is, setting the specification to a limit of a predicted genotoxic impurity in the drug substances to a value of 1.5 μg/day/person.

2.2.3 Prediction of Phospholipidosis

Phospholipidosis (PLP) is not a toxic effect *per se*, but rather an adaptive histological alteration, caused either by direct inhibition of the phospholipid degradation pathway (i.e., phospholipase inhibition) or formation of complexes between chemicals and phospholipids, which block the access of phospholipase. Depending on the extent and duration, PLP may cause secondary tissue damage or functional disorders. Drug-induced PLP is described for some approved pharmaceuticals. It is debatable whether the histological alteration *per se* is harmful and the correlation between preclinical *in vivo* results and the human effects is questionable [12]. On the other hand, the chemical properties that lead to PLP are well understood and can be derived directly from the structural formula [13]. The majority of PLP-causing compounds are cationic amphiphiles, meaning that besides carrying a positive charge, they have both a lipophilic and a hydrophilic structural component. These properties are clearly related to the mechanism: their amphiphilic nature will allow them to migrate through the lipid bilayer of biomembranes and the cationic charge will lead to trapping in lysosomes that have an acidic pH. The available *in silico* models generally show a reasonable predictivity for the *in vivo* situation, that is, occurrence of foamy macrophages and/or confirmation of PLP in electron micrographs [14]. In addition, the predicted results show a high concordance with *in vitro* assays, allowing to gradually eliminate the PLP properties in chemical series already in early drug development.

Such a strategy lowers the chance for critical preclinical *in vivo* findings that would necessitate additional mechanistic investigations or cumbersome monitoring during clinical trials, which is the reason why many pharmaceutical companies have such predictive tools in place even without guideline requirement or regulatory recommendations.

2.2.4 Prediction of Carcinogenicity

The concept of structural alerts is also implemented for carcinogenicity prediction. The broadest data source for assessing carcinogenicity stems from the US National Toxicology Program and consists of the results of the rodent carcinogenicity bioassay. Chemicals that showed carcinogenic effects in the animal studies were analyzed for structural commonalities from which structural alerts were derived [15]. These alerts together with further literature results are implemented in several predictive systems (DEREK Nexus, OncoLogic, ToxTree, HazardExpert, and MultiCase). ToxTree has included a refinement of the structural alerts by integrating local quantitative models for aromatic amines, which often deliver ambiguous results, if the prediction is exclusively based on structural alerts [16].

All these systems identify a hazard, which is helpful in situations where an exposure of a potentially carcinogenic compound has to be strictly avoided because accidental exposure has no benefit for the consumer. In the field of pharmaceuticals, however, the situation is different: depending on the benefit for the patient, a carcinogenic hazard is not necessarily prohibitive (this is particularly true for oncology drugs), but it is important to know the potency of the carcinogenic effect and the expected exposure. In addition, the difference between genotoxic carcinogens, for which a no-effect threshold is difficult to be established, and nongenotoxic carcinogens, which clearly have thresholds below which an effect is unlikely to occur, is important for risk assessment. Whereas genotoxic carcinogens are already identified with the above-mentioned mutagenicity prediction tools, the prediction of a nongenotoxic carcinogenic hazard might perhaps be helpful for deriving mechanistic hypotheses in cases where proliferation and/or hyperplasia are observed in animal studies, but it will most probably not replace the *in vivo* risk assessment for drugs in the near future.

2.2.5 Prediction of Skin Sensitization

Sensitization of skin and lung due to exposure of chemicals is mainly due to reactions of chemical (sub)structures with proteins, which in turn may elicit the formation of antibodies, a process called haptenization. Besides the capability for forming protein conjugates, skin resorption and skin metabolism are further important parameters that affect the allergic potential of a compound. Skin resorption is mainly influenced by the physicochemical

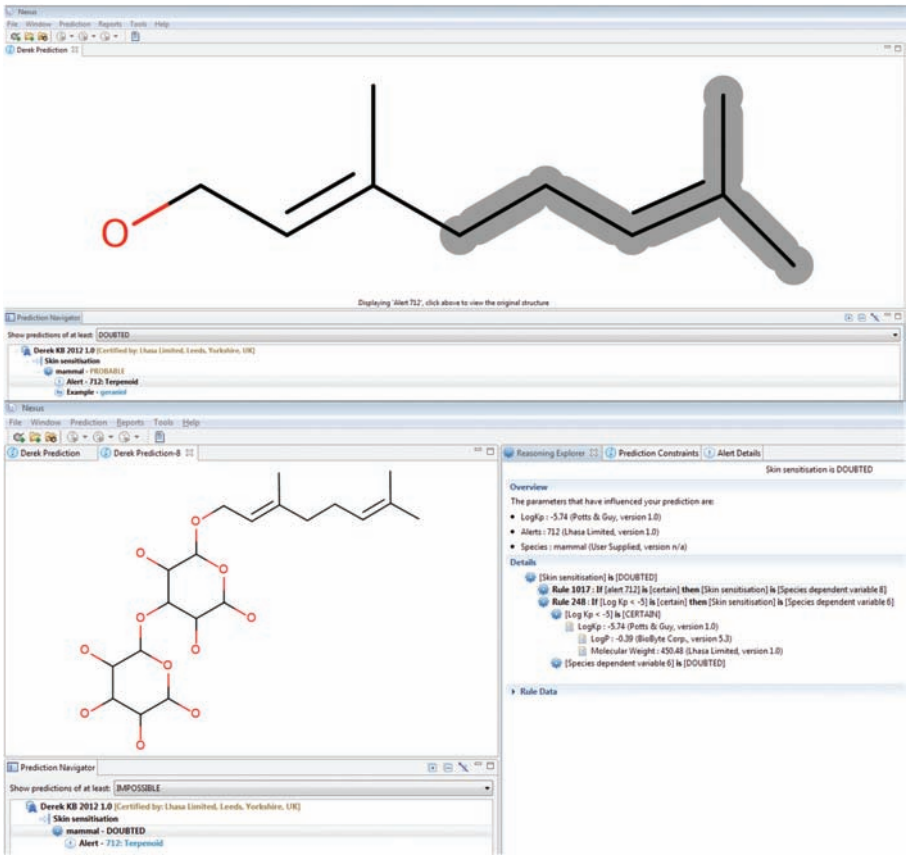


Figure 2.1 An illustration of how a combination of structural alerts with predicted physicochemical data may influence prediction for skin sensitization. Geraniol is a known skin sensitizer and consequently triggers the structural alert for terpenoids in DEREK Nexus (Version 3.0.1). If a sugar backbone is added to geraniol, the terpenoid structure is still recognized but due to the fact that the calculated permeability log K_p is below −5 this hypothetical structure is assumed to have low skin permeability and thus the overall prediction for skin sensitization is set to "doubted."

properties, such as size and fat solubility (lipophilicity) of the molecules. Through the combination of algorithms predicting both reactivity and physicochemical properties, reliable predictions may be obtained for skin sensitization (Figure 2.1) [17].

In a review of existing *in silico* tools for the prediction of allergic contact sensitization, the authors conclude that the models that are based on mechanistic models generally perform better than those that are based on purely statistical models. Independently of the approach, the potency of a sensitizer is inadequately

predicted by all modeling approaches [18], that is, again the tools are limited to hazard identification.

2.2.6 Prediction of Skin and Eye Irritation

For the assessment of skin irritation, the analysis of structure–activity relationship is recommended in the relevant experimental testing guidelines before embarking into an *in vivo* study [19]. First rule-based approaches to identify skin and eye irritants were implemented in the *in silico* decision support system by the German Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV) [20]. The rules were subsequently included and advanced in the open source software ToxTree [21]. The sensitivity of such systems, that is, the power to identify an irritant or corrosive compound, is considered to be acceptable. However, it is also clearly stated in the quoted guideline that "negative data from studies of structurally related substances or mixtures of such substances do not constitute sufficient evidence of noncorrosivity/nonirritancy of a substance under the sequential testing strategy," which will consequently result in experimental testing. The example of *in silico* prediction of skin irritation reveals some general principles for *in silico* tools. The guideline sets three prerequisites for a meaningful use of an SAR tool, which can be generalized for other endpoints, too:

1) Availability of sufficient human and/or animal data.
2) Structural relationship of the substances for which data are available with the compound under question.
3) Indication of corrosion/irritancy *potential*.

It is certainly debatable at what point prerequisite no. 1 is fulfilled, that is, when is the required data "sufficient." For mutagenicity data or, more strictly speaking, for data from the *Salmonella* reverse mutation assay, there are certainly abundant data sets in the public domain. The number of data sets for skin irritation tests (usually the Draize test according to OECD Guideline No. 404) is significantly lower. At the end, the answer to the question of how much data are needed depends mainly on the diversity of the chemical series that have to be predicted. This leads to prerequisite no. 2, that is, structural relationship. The first step to assess a structural relationship is certainly the expert's eye, but tools have been developed to assist in the task to determine similarity between the test compound and the compound to be predicted. Similarity is not limited to a comparison of the graphical representation of the chemical, but implies that "structurally similar molecules are expected to exhibit similar properties or similar biological activities" [22]. It is beyond the scope of this chapter to summarize the different approaches for assessing structural relationship. A look into the structure and the identification of substructures will probably suffice for mechanistically simple endpoints, such as mutagenicity or skin irritation. For these endpoints, the

biological effect is related to chemical reactivity, which in turn can be derived from the structural components. For endpoints that have slightly more complex mechanism, for example, micronucleus formation, this approach might be too short sighted. Micronucleus formation can be caused by direct DNA cleavage (clastogenic effect), but it may also be caused by impairment of the spindle fiber apparatus. This in turn may be elicited by influences on tubulin polymerization/depolymerization or the attachment of the spindles to the kinetochores. These different mechanisms may be caused by completely different (sub)structures and will only be depicted if effect data sets are correlated with several chemical descriptors and not just the analysis of similarity of the graphical structure.

The word "potential" is not used in an unambiguous way, but it often implies some level of quantitative information for the biological effect. As an example, for dermal irritation or corrosion this is reflected in the grading of the observed skin reaction as outlined in the guideline [19]. However, this information, only available in the original reports, is not always available for the construction of *in silico* systems. In some cases, chemicals tested for specific endpoints are binned in a binary or tertiary system with "yes and no" or "no, mild, or severe." In contrast to the original data, this information can be derived from the Globally Harmonized System classification (which replaces the former R-phrases). The chemicals in each bin can then be searched for structural commonalities and the derived alerts can be programmed into a computer system [23]. Toxicity studies for hazard classification are performed according to guidelines, which often require application of fixed doses or concentrations that have to be applied in the assays (e.g., for eye irritation studies, the dose is set to 0.5 ml of liquid or 0.5 g of solid or paste that is to be applied to the test site; for acute oral toxicity classification, animals are dosed in a stepwise procedure using the fixed doses of 5, 50, 300, and 2000 mg/kg). As a consequence, there is inherently some level of exposure information in the final classification result, even though internal exposure, that is, plasma or organ levels, is normally not determined, but without access to the original data of the study, this potency information may get lost in the binning approach described above; that is, potency prediction is currently underdeveloped in most of the available systems.

2.2.7 Approaches to Systemic Toxicity Prediction

2.2.7.1 The Cramer Classes

The Cramer classes represent a system to rank chemicals of unknown toxicity into three toxicity classes (low, moderate, and serious). It is intended for deciding on the priority of toxicity investigation and thus still plays an important role in the context of safety evaluation of industrial chemicals that have not already been notified or regulated. The approach dates back to 1978 and can be seen as a prototype of an expert system [24]. It is based on a decision tree, where questions mainly related to chemical reactivity are posed, for example, "Does the structure contain elements other than carbon, hydrogen, oxygen, nitrogen, or

divalent sulfur?" The answers are provided in a binary way based on the structural constituents of the compound under question. The Cramer decision tree and the structural alerts together with some refinements are implemented in ToxTree and the OECD QSAR toolbox. In the original publication, Cramer and coworkers correlated the three classes with no observed adverse effect levels (NOAELs) of investigated compounds falling into each class. From this correlation, they derived a "protection index" (PI) based on daily per capita intake of consumers. Even though this is an approach to derive risk estimated from hazard identification via NOAELs, the high level of uncertainty included into the PIs prevented a widespread use of this procedure in risk assessment of chemicals.

2.2.7.2 Predicting Toxic Doses of Drugs

An alternative approach to predict systemic toxicity was followed by Contrera *et al.* [25] to predict human toxicity of new drug candidates. From prescription information of marketed drugs, the authors collected the values for the maximum recommended daily dose (MRDD, provided as mg/kg/day), assuming that this dose limits the therapeutic window. Consequently, toxic effects have to be expected if it is surpassed. The chemical structures representing the drugs were then analyzed for occurrence of a broad variety of descriptors and these in turn correlated with the MRDD values. The reported concordance, specificity, and sensitivity values for the resulting QSAR are all above 75%, thus indicating a reliable statistical correlation. The drawback of this black-box approach though is that it does not allow for a conclusion on the mechanism of a potential effect, that is, is it caused by an off-target toxicity of the drug candidate or is toxicity due to excessive pharmacodynamics? However, information of a potential mechanism of toxicity and the affected organ or tissue are important aspects in order to assess how well a potential toxicity can be monitored in a clinical trial. In addition, the QSAR does not directly provide a prediction of a no-effect level. The latter is important to derive a safe starting dose for the first-in-human clinical trials during drug development, which is currently still based on the results of the preceding animal studies. Consequently, the system has not found routine entry into drug candidate evaluation before the first clinical trial.

2.2.7.3 Predicting Organ Toxicity

Theoretically, the techniques described in Section 2.2.2 can also be applied to develop models for organ toxicity; that is, descriptors for the chemical structures that cause liver, heart, or kidney toxicity can be correlated with their toxicological effect. The caveat here is that the observed toxicity can be caused by a multitude of individual mechanistic steps, and each of them can be affected by the molecule under consideration or its metabolites. As a consequence, models based only on chemical descriptors usually show poor performance in the prediction of organ toxicity, as demonstrated for the prediction of drug-induced liver toxicity [26]. However, if the correlation is based on existing *in vivo* toxicity data, the mere chemical approach can be refined by adding additional results from these systemic studies in order to better represent the whole system.

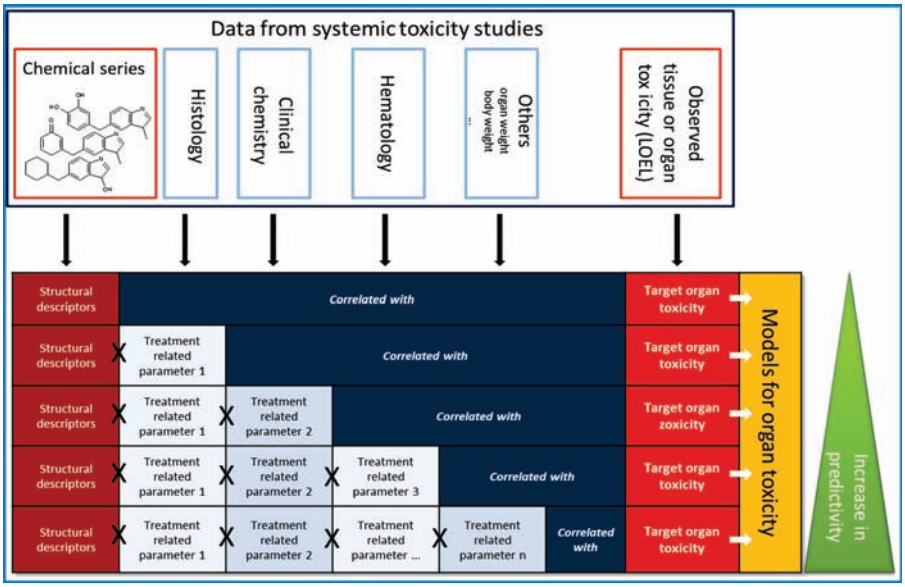


Figure 2.2 A depiction of how data from systemic *in vivo* studies could be used to build models for predicting organ toxicity. It is expected that the predictivity of such models increases when more endpoints are used to build the models due to an increase in weight of evidence.

Essentially, the selection of the parameters to be added can be done in the same way as a toxicologist would assess organ toxicity based on the results of an *in vivo* study. For example, if the liver of an animal is enlarged or shows a weight increase, this represents a first hint for liver toxicity. The enlargement could as well be due to the induction of metabolizing enzymes. To assess the weight of evidence of liver damage, additional data are therefore required, primarily the histopathological assessment (signs of inflammation, fibrosis, steatosis, cholestasis, hyperplasia, etc.). Histopathology is still the gold standard for organ toxicity evaluation. In addition, further parameters may be considered for constructing models, such as elevation of transaminases, increase of total bilirubin, and changes in blood coagulation parameters, to increase the weight of evidence and thus predictivity as depicted in Figure 2.2.

The advantage of this approach lies in the fact that assessment criteria, which are well established in the field of toxicology, such as "an elevation of transaminase activities (alanine aminotransferase and/or aspartate aminotransferase) higher than threefold the upper limit of normal accompanied by an at least twofold increase above the upper limit of normal total bilirubin concentration may indicate functionally significant liver damage," can be implemented into such a system [27]. However, the fundamental drawback is the lack of abundant well-curated, harmonized data. Results of liver transaminase values are not easily available in the public domain and the terminology for these enzymes is not

always harmonized, not to speak about histopathology, where standardized terms are even less common. These deficits are still one of the main reasons why *in silico* systems based on preclinical *in vivo* toxicity data are currently not available.

2.2.7.4 Adverse Outcome Pathways and Potential for Prediction

The search for alternatives to animal testing and the limitations of the purely descriptive classical toxicological study designs triggered the evolution of the adverse outcome pathway (AOP) analyses. The idea behind AOP is to dissect the observed toxic effect into individual steps or sequences that can each be assigned to a specific mechanistic event as displayed in Figure 2.3.

The advantage of this procedure is that the individual steps are amenable to *in silico* or *in vitro* experimental assessment and might therefore replace or reduce animal experimentation, if they are fully understood and elucidated. The implementation of an adverse outcome pathway for a specific toxicity is, however, not trivial, especially if the pathway leading to the organ response is branched or also affected by metabolites. For chemically induced skin sensitization, a dissection of the observed effect into individual steps has been proposed as displayed in Figure 2.4 [29].

The dissection illustrates which part of the pathway can be investigated by *in silico* or *in vitro* models; for example, the electrophilic nature of a compound can be predicted by *in silico* tools, whereas activation of dendritic cells may be tested by *in vitro* assays. As soon as sufficient experimental data become available for the individual biological descriptors, these can in turn be modeled *in silico*. This may eventually result in a multilevel model (see Figure 2.5), where the individual results are integrated into an overall organ toxicity prediction by either sophisticated reasoning or multiparameter statistical approaches. Skin sensitization is currently probably the most advanced AOP due to the needs of the cosmetic industry; however, there is rapid progress also for other endpoints, such as cholestasis, liver fibrosis, and steatosis [30].

The *in silico* approach depicted in Figure 2.5 currently exists only on a conceptual level. However, there is first evidence that the way forward is feasible as recently shown for the endpoint drug-induced liver toxicity: the combination of chemical

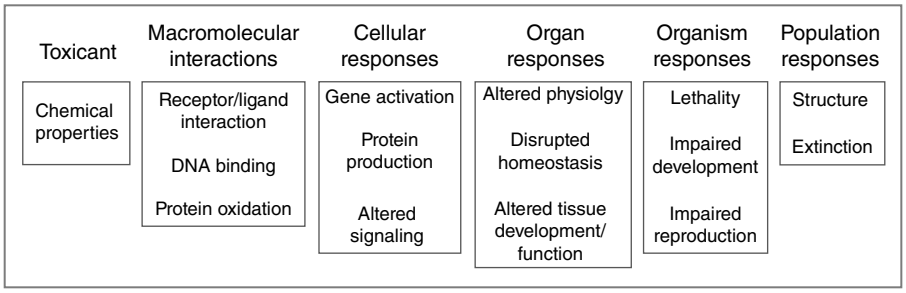


Figure 2.3 The adverse outcome pathway concept according to OECD [28]. For the purpose of animal or human toxicity analysis, the population response (the last box on the right end) can be ignored.

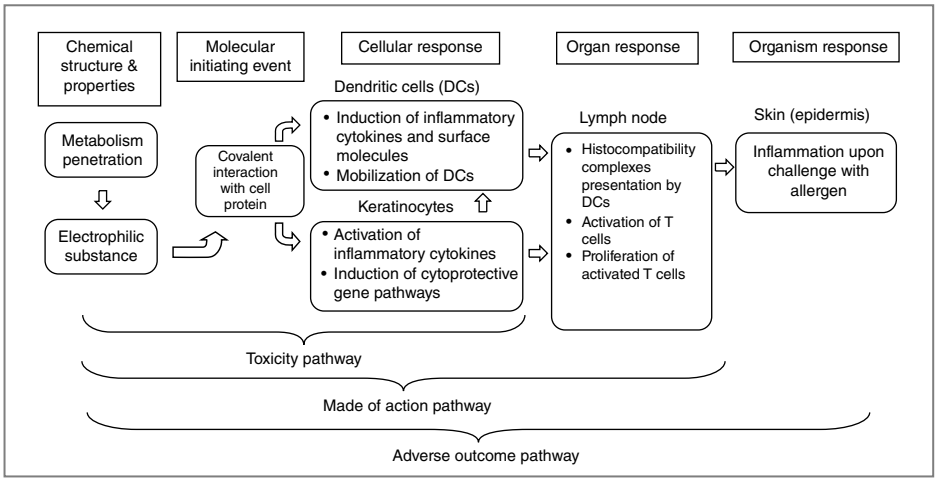


Figure 2.4 The adverse outcome pathway for skin sensitization according to OECD [29].

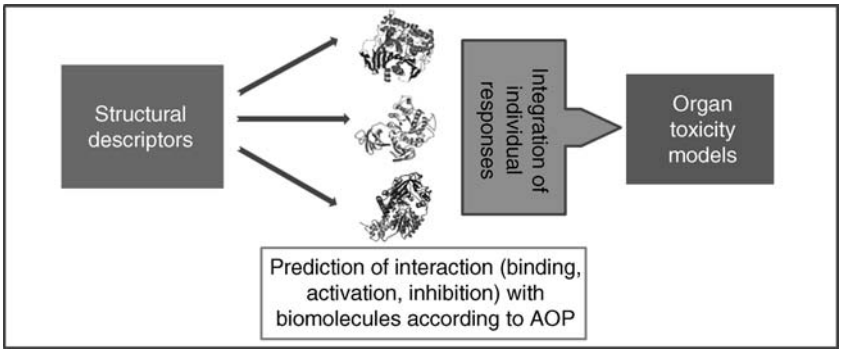


Figure 2.5 Strategy for *in silico* modeling based on the adverse outcome pathway.

descriptors with data from hepatotoxicity-specific cellular responses obtained with high-content imaging clearly outperformed the pure chemistry approach in predicting the liver effects [26]. As already stated above, availability of sufficient experimental data is still the major hurdle for rapid progression of this concept.

2.3 Prediction of Risk

2.3.1 Risk Definition and Some Basic Considerations

The United States Nuclear Regulatory Commission provides the following definition for risk: "The combined answer to three questions that consider (1) what

can go wrong, (2) how likely it is, and (3) what its consequences might be" [31]. Applied to toxicological risk assessment "what can go wrong" is the "hazard" as defined further above, whereas the likelihood and the consequences of an effect are determined by the level of exposure of the animal or the human being for which the risk assessment is to be performed. In a simplified version, the relationship can be described by the following formula:

$$\text{Risk} = \text{Hazard} \times \text{Exposure}.$$

In Section 2.2, it has been described how *in silico* tools can be built including some level of potency or exposure information to progress from hazard to risk prediction. Defining a threshold below which no toxic effect is expected represents such an approach. If a TTC is determined, one can back-calculate the maximum dose or exposure that would be acceptable in order to stay below this threshold. Alternatively, the doses in the experimental studies are kept fixed, thus assuming a fixed range of exposures, in order to derive toxic classes of compounds. However, all these approaches have shortcomings. The TTC concept can estimate a threshold risk but is barely able to quantify the risk if exposure occurs above the threshold, because it contains no information on the steepness of the dose–response curve. The fixed dose approach to derive toxic classes in the OECD studies is able to differentiate between very toxic and less innocuous compounds, but will not predict the level of effect at a specific exposure. These shortcomings become particularly obvious if one considers the design of toxicological studies performed during drug development. These studies have the main objective to determine the safe first-in-human dose for clinical trials. The systemic toxicity studies are designed in a way that they identify key thresholds such as the "no observed effect level" (NOEL), the NOAEL, the lowest observed effect level (LOEL), and the maximum tolerated dose (MTD). These thresholds are subsequently used for deriving the first safe dose in humans. The doses administered in these studies are selected on the basis of available *in vivo* pharmacological results, previous dose range-finding studies, considerations of bioavailability, expected human therapeutic dose or exposure, and feasibility of maximum application volume. The thresholds are determined by the observations made for the different endpoints investigated in these studies, that is, clinical observations (behavior, body weight, food and water consumption), clinical chemistry, hematology, blood hemostasis, gross morphology, and histopathology. All of these endpoints also including histopathological findings have a quantitative component (severity, grading, number of animals affected). Observed changes in an endpoint are considered to be compound-related, if the following criteria are met:

- the changes show a dose–effect relationship (i.e., occur at a low dose and increase with augmenting dose),
- the changes are significantly different from the vehicle or control group,
- the size of the change is of biological relevance (some changes might be statistically significant, but too small to be of biological relevance),

- the changes pass a certain level or magnitude (e.g., twofold elevation of liver transaminases) or the changes rise above levels of historical control data.

The assessment of observations and the determination of effect levels are always based on expert judgment, which usually involves several toxicological disciplines. Although the effect levels are generally expressed as administered dose per body weight, systemic toxicity studies usually also harness toxicokinetic information, that is, the internal exposure (plasma or serum levels). For the calculation of the safe human starting dose, the definition of an NOAEL is in most cases based on an effect that would be unacceptable if produced by the initial dose of a therapeutic in a phase I clinical trial conducted in adult healthy volunteers [32].

It is clear from the procedure described above that a binary or tertiary binning or identification of different toxic classes as used for hazard assessment is not feasible for predicting risk from *in vivo* toxicity data. The data are far too heterogeneous for such a simplified approach. As a consequence, there are currently no *in silico* systems available, which predict the risk of adverse effects for humans or animals upon exposure to an unknown compound, based on data derived from *in vivo* toxicity data.

2.3.2 Data Availability

Shortcomings in computational power or lack of mathematical procedures are not the main reason for the lack of such *in vivo* prediction system. The key hurdle is certainly data availability. *In vivo* toxicity data are published since decades in scientific journals, but the difficulty is to bring these data sets into a meaningful structure. In addition, the endpoints for which the data are collected need some level of comparability; that is, the methods that were used to, for example, measure transaminases in clinical chemistry need to be known and ideally be performed according to harmonized protocols. This demand has long been raised and is not unique for *in silico* tools. As a consequence, numerous guidelines have been implemented for harmonized performance of *in vivo* studies (e.g., OECD, ICH, NTP). Available databases such as the National Toxicology Program [33] contain abundant toxicity data, but the query functions for structures or specific endpoints are often limited. Such limitations and restrictions of the publicly or commercially available toxicological databases were analyzed by the Structure–Activity Relationship Database Project Committee under the auspices of the International Life Sciences Institute (ILSI) in 2005. To improve data availability and retrieval, the committee proposed a pilot for toxicology information source with the name International Toxicology Information Centre [34]. The pilot was subsequently populated with data from NTP, the EPA Gene-Tox database, the Carcinogenicity Potency Database ("Gold" database) [35], and IUCLID (European Chemical Notification database) [36]. The mentioned data sets were selected because they are publicly available and reported at a level

of granularity that is close to the original observed data performed according to internationally agreed guidelines. A similar approach was undertaken by the Fraunhofer Institute of Toxicology and Experimental Medicine in their database RepDose [37]. RepDose currently covers around 2300 subacute to chronic toxicity studies, performed mainly on industrial chemicals in rats, mice, or dogs with oral or inhalation exposure. The field of pharmaceuticals has been less well covered in both databases, mainly due to the fact that the majority of the toxicological data of drug candidates that failed during development are usually not published. The toxicological data for approved drugs are only available as summary in public assessment reports but not in databases. To overcome this shortage of searchable toxicology data on pharmaceutical compounds, the European pharmaceutical industry organized in EFPIA (European Federation of Pharmaceutical Industries and Associations) together with the European Innovative Medicines Initiative founded the eTOX project ("electronic toxicity"). Thirteen pharmaceutical companies started to share preclinical safety data and to further develop a database for read-across, data mining, and modeling [38]. Besides setting up a modified database and work streams for safe data sharing (more than 5000 reports on systemic toxicities studies have already found their entry into the eTOX database), the main achievement of the project is a series of ontologies for the different endpoints investigated in *in vivo* studies. Even though most of the studies are performed according to harmonized protocols, reporting usually occurs on an individual basis, making comparison of findings between companies rather difficult. For example, for one and the same finding in histopathology different terms may be reported in different companies. Therefore, such ontologies are key for usability and interoperability of databases. Ontology development in eTOX went hand in hand with similar approaches of FDA originating for clinical data. The Clinical Data Interchange Standards Consortium (CDISC) [39] has extended its activity toward preclinical data, the activity being called Standard for Exchange of Nonclinical Data (SEND). The developed standards will soon become mandatory for preclinical data exchange with FDA during IND and dossier submission.

eTOX is not the only initiative collecting toxicity data and making it accessible for read-across and creation of new *in silico* tools. Another example is the Japanese Hazard Evaluation Support System (HESS) that was initiated by the Japanese Ministry of Economy under participation of the National Institute of Health Science focusing rather on industrial chemicals but with the same objectives to gather data and create integrated platforms for the prediction of risk of new chemicals [40]. Aligning the different initiatives will certainly be a valuable objective for the near future.

2.3.3 Database Structure and Data Curation

Two main issues have to be resolved during the construction of a database. First, a decision on the level of data granularity has to be made. Second, the collected

results have to be curated in order to harmonize differences in the description of findings. The first issue is mainly influenced by the data sources: if only summaries of toxicological data are available for the construction of the database, the depth of data will not reach very deep. If, however, the original data of a regulatory 4-week toxicity data are the starting point, the data can be extended to individual values for body weight, food consumption, transaminase elevation, and so on. Since such studies may consist of data sets of several hundred pages, a pragmatic limit has to be set in order to allow data extraction from these reports within meaningful time and with reasonable effort. Ideally, the database will be structured in such a way that a query will result in answers a toxicologist would raise for a new compound:

- Which target organ was affected?
- At what dose did the effect first occur?
- Was the effect treatment related or a chance finding?
- Was there a correlation between organ weight changes, histopathology, and organ-specific biomarkers (e.g., transaminases)?

These questions determine not only the complexity of the data collection but also the required query functions of the database. Searching for structures including substructure or similarity search is mandatory but not sufficient to accommodate the needs of complex queries. Complex queries need to combine structural features with individual findings connecting via ontologies allowing also to differentiate between "treatment-related findings" and "non-treatment-related findings." The latter is an important aspect in the population and curation of a database: the original findings in a systemic toxicity study are usually assessed by a team of experts, which carefully decides whether a finding is to be attributed to the compound under investigation or whether it is a chance finding. However, chance findings should not be omitted from the database because when mining large data sets, chance findings might also reveal interesting trends. Figure 2.6 provides an example for a search in the eTOX database for compounds carrying a diphenylmethyl substructure, which also show treatment-related cellular lipid accumulation vacuolation, an effect that may be related to phospholipidosis.

The result table illustrates how important the curation of the collected data is. The finding "foamy macrophages" is a description of what is seen during lipid accumulation in the context of drug-induced phospholipidosis. It could also be described as lipid accumulation, phospholipidosis, or something similar. Without any ontology, which relates these different terms, a database with this level of detail would be almost worthless and certainly not helpful for data mining. The need for ontologies relates to all endpoints covered in the database. Standard nomenclature exists for the transaminases measured in clinical pathology; however, for histopathology these solutions have just started to emerge. Regulatory agencies have also identified the need of such standardized terms and therefore extend their existing activities in the field of harmonizing clinical data into the development of preclinical ontologies [39].

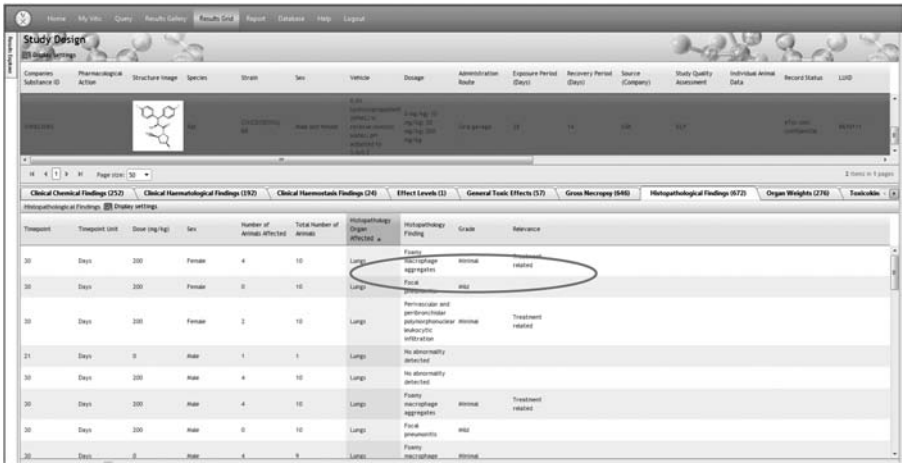


Figure 2.6 Screenshot of the eTOX database (database version v2014.1) result table as viewed through VITIC Nexus (Version 2.5.0, provided by Lhasa Ltd.) for a complex query including a structural component, the search of a specific histopathological finding, and the question whether the finding is treatment related or incidental. The query results in a compound for which foamy macrophage aggregates were observed in the high-dose group (ellipse) in 4 out of 10 animals. The finding was considered to be treatment related.

The advent of large toxicological databases with rather detailed result recording together with ontologies for the recorded endpoints in place should greatly facilitate complex searches for read-across purposes or data mining. It could even be envisaged that such searches reveal underlying relationships not yet detected, because the previously existing data sets were too small and therefore the effects did not surpass the signal-to-noise ratio.

2.3.4 Approaches to Model and Predict Risk

One of the headers for data section in Figure 2.6 lists toxicokinetics, that is, the database includes not only the dose information but also serum or plasma levels. Thus, the database fulfills one key prerequisite to model risk, which is the exposure information. Unfortunately, approaches to build predictive systems using this information are still in their infancy and it would be premature to provide an overview here. Generally, the idea of how to combine the hazard prediction with exposure assessment is to include additional tools that predict exposure based on physiologically based pharmacokinetic (PBPK) models. PBPK models are quantitative descriptions of the absorption, distribution, metabolism, and excretion (ADME) [41], where each predicted ADME parameter relies on different physicochemical or biochemical descriptors. For example, prediction of absorption could solely rely on the physicochemical characterization of a chemical, but it could also include first *in vitro* permeability assessment in colon carcinoma cell lines (CaCo-2 model). PBPK models are not restricted to the

prediction of serum or plasma exposure, but rather they can predict exposure of individual organs or tissues. Such an exposure is usually not assessed in systemic toxicity studies. Thus, these models bear great potential in the context of the above-described adverse outcome pathways, where effects on individual cells or tissues may be identified as causative trigger for organ toxicity. However, up to now PBPK models are mainly used to predict therapeutic human exposure, for example, for the planning of a first-in-human study during drug development. The combination of such models with hazard prediction still exists only on a conceptual frame [42]. It remains to be seen whether the broad availability of solid and well-curated *in vivo* data will result in meaningful models for risk prediction.

2.4 Thoughts on Validation

With a more widespread use of *in silico* tools, discussion started on how to validate such systems in order increase the level of certainty regarding reliability of prediction and applicability of the tools. In 2007, the OECD published a guidance document setting some minimum requirements for validation [43]. According to the guidance, the cornerstone of validation is provision of information on the following aspects of an *in silico* model:

- Definition of the predicted endpoint.
- Unambiguous algorithm.
- Appropriate measures of goodness-of-fit, robustness, and predictivity.
- Definition of the domain of applicability.
- Mechanistic interpretation, if possible.

Whereas the fulfillment of these requirements is still conceivable for a model with a simple endpoint such as mutagenicity, there are currently no validated models for more complex endpoints. Broad consensus exists about the determination of goodness-of-fit (how well does the model perform on compounds it was trained with) and the measures for the external predictivity (how well does the model perform on compounds that were not included in the training set). The most common parameters are summarized in Table 2.1.

The applicability domain defines the area of chemical space inside which it can reliably make predictions. It is generally determined by the chemical space of the model training set. The procedures for defining the applicability domain are less well agreed upon. The most common approach is the definition directly from the chemical structures of the training series, but other procedures were proposed that may be more meaningful depending on the nature of the model [44]. Validation of models must always be seen as a coevolutionary process with model development. There are currently no fully validated models for regulatory purposes, but it is clear that a maximum of transparency and documentation will provide the best foundation for a validation.

Table 2.1 Summary of statistical measures of goodness-of-fit (predictivity for compounds used for building and training the model) and for description of external predictivity (i.e., prediction for compounds not included in the training set).

Endpoint type	Statistical measure
Quantitative	R^2 – determination coefficient/coefficient of variation
	RMSE – root mean square error
	Q^2_{LOO} – leave-one-out (LOO) cross-validated determination coefficient
	SDEP – standard error of prediction
Qualitative (usually binary, for example, mutagenic/nonmutagenic)	TP – true positive
	TN – true negative
	FP – false positive
	FN – false negative
	Specificity
	Sensitivity
	MCC – Matthew's correlation coefficient

2.5 Conclusions and Outlook

Currently, reliable *in silico* models are available for only a few endpoints and robust prediction of *in vivo* toxicity or adverse events affecting specific organs are not at hand. Whereas decades ago the application of complex algorithms was limited by the available computer power, this hurdle has now been overcome. The limiting factor for modeling organ or *in vivo* toxicity is currently access to abundant, high-quality, and well-curated data. Several consortia have been formed to overcome this limitation and it is hoped that the consortia and data sets will merge in the future. Data availability *per se* is already a major asset for read-across approaches, but it is also expected that new *in silico* systems and tools will be developed based on improved data accessibility. This will also apply to endpoints that have not been discussed in this chapter because *in silico* approaches are scarcely developed (e.g., for reproductive toxicity). In addition, it is warranted to anticipate significant contributions of these new systems to animal welfare and 3R (reduce, refine, and replace). The rise of *in silico* toxicology has just begun.

References

1 Steger-Hartmann, T. and Scott, B. (2014) Chapter 36: Computer-based prediction models in regulatory toxicology, in *Handbook of Regulatory Toxicology* (eds F.-X. Reichl and M. Schwenk), Springer, Heidelberg.

2 Simon-Hettich, B., Rothfuss, A., and Steger-Hartmann, T. (2006) Use of computer-assisted prediction of toxic effects of chemical substances. *Toxicology*, **224**, 156–162.

3 Steger-Hartmann, T. (2013) Guest editorial: advances in computational toxicology. *Mol. Inform.*, **32** (1), 9 (special issue).

4 Faustman, E.M. and Omenn, G.S. (2013) Risk assessment, in *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 8th edn (ed. C.D. Klaassen), McGraw-Hill Medical Publishing Division, New York, pp. 123–125.

5 Ashby, J. and Tennant, R.W. (1988) Chemical structure, *Salmonella* mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat. Res.*, **204**, 17–115.

6 Knudson, A.G., Hethcote, H.W., and Brown, B.W. (1975) Mutation and childhood cancer: a probabilistic model for the incidence of retinoblastoma. *Proc. Natl. Acad. Sci. USA*, **72**, 5116–5120.

7 Müller, L., Mauthe, R.J., Riley, C.M., Andino, M.M., De Antonis, D., Beels, C., DeGeorge, J., De Knaep, A.G.M., Ellison, D., Fagerland, J.A., Frank, R., Fritschel, B., Galloway, S., Harpur, E., Humfrey, C.D.N., Jacks, A.S.J., Jagota, N., Mackinnon, J., Mohan, G., Ness, D.K., O'Donovan, M.R., Smith, M.D., Vudathala, G., and Yotti, L. (2006) A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. *Regul. Toxicol. Pharmacol.*, **44**, 198–211.

8 Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., and Würtzen, G. (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. *Food Chem. Toxicol.*, **42**, 65–83.

9 Hansen, K., Mika, S., Schroeter, T., Sutter, A., ter Laak, A., Steger-Hartmann, T., Heinrich, N., and Muller, K.R. (2009) Benchmark data set for *in silico* prediction of Ames mutagenicity. *J. Chem. Inf. Model.*, **49**, 2077–2081.

10 International Conference on Harmonisation (2013) Genotoxic impurities M7: assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. EMA/CHMP/ICH/83812/2013.

11 Sutter, A., Amberg, A., Boyer, S., Brigo, A., Contrera, J.F., Custer, L.L., Dobo, K.L., Gervais, V., Glowienke, S., Gompel, J.V., Greene, N., Muster, W., Nicolette, J., Reddy, M.V., Thybaud, V., Vock, E., White, A.T., and Müller, L. (2013) Use of *in silico* systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. *Regul. Toxicol. Pharmacol.*, **67** (1), 39–52.

12 Greaves, P. (2012) *Histopathology of Preclinical Toxicity Studies: Interpretation and Relevance in Drug Safety Evaluation*, 4th edn, Academic Press, London.

13 Hanumegowda, U.M., Wenke, G., Regueiro-Ren, A., Yordanova, R., Corradi, J.P., and Adams, S.P. (2010) Phospholipidosis as a function of basicity, lipophilicity, and volume of distribution of compounds. *Chem. Res. Toxicol.*, **23**, 749–755.

14 Orogo, A.M., Choi, S.S., Minnier, B.L., and Kruhlak, N.L. (2012) Construction and consensus performance of (Q)SAR models for predicting phospholipidosis using a dataset of 743 compounds. *Mol. Inform.*, **31** (19), 725–739.

15 Benigni, R. and Zito, R. (2004) The second national toxicology program comparative exercise on the prediction of rodent carcinogenicity: definitive results. *Mutat. Res.*, **566**, 49–63.

16 Benigni, R., Bossa, C., Jeliazkova, N., Netzeva, T., and Worth, A. (2008) The Benigni/Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. EUR 23241 EN. Available at http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/EUR_23241_EN.pdf (last accessed June 20, 2014).

17 Dimitrov, S.D., Low, L.K., Patlewicz, G.Y., Kern, P.S., Dimitrova, G.D., Comber, M. H., Niemla, J., Bailey, P.T., and Mekenyan, O.G. (2005) Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int. J. Toxicol.*, **24**, 189–204.

18 Teubner, W., Mehling, A., Schuster, P.X., Guth, K., Worth, A., Burton, J., van Ravenzwaay, B., and Landsiedel, R. (2013) Computer models versus reality: how well do *in silico* models currently predict the

sensitization of a substance. *Regul. Toxicol. Pharmacol.*, **67** (3), 468–485.
19 Organisation for Economic Co-operation and Development (2002) OECD Guideline for the Testing of Chemicals: Acute Dermal Irritation/Corrosion. OECD/OCDE 404.
20 Gerner, I., Zinke, S., Graetschel, G., and Schlede, E. (2000) Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Creation of fundamental rules for a decision support system. *Altern. Lab. Anim.*, **28** (5), 665–698.
21 Joint Research Centre, Institute for Health and Consumer Protection (2013) ToxTree. Available at http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/toxtree (last accessed June 20, 2014).
22 Kochev, N., Monev, V., and Bangov, I. (2003) Searching chemical structures, in *Chemoinformatics – A Textbook* (eds J. Gasteiger and T. Engel), John Wiley & Sons, Ltd, Chichester, pp. 291–316.
23 Gerner, I., Liebsch, M., and Spielmann, H. (2005) Assessment of the eye irritating properties of chemicals by applying alternatives to the Draize rabbit eye test: the use of QSARs and *in vitro* tests for the classification of eye irritation. *Altern. Lab. Anim.*, **33**, 215–237.
24 Cramer, G.M., Ford, R.A., and Hall, R.L. (1976) Estimation of toxic hazard – decision tree approach. *Food Cosmet. Toxicol.*, **16**, 255–276.
25 Contrera, J.F., Matthews, E.J., Kruhlak, N. L., and Benz, R.D. (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. *Regul. Toxicol. Pharmacol.*, **40**, 185–206.
26 Zhu, X.W., Sedykh, A., and Liu, S.S. (2014) Hybrid *in silico* models for drug-induced liver injury using chemical descriptors and *in vitro* cell-imaging information. *J. Appl. Toxicol.*, **34** (3), 281–288.
27 Food and Drug Administration (2009) Guidance for Industry: Drug-Induced Liver Injury: Premarketing Clinical Evaluation. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER).
28 Organisation for Economic Co-operation and Development (2013) Adverse Outcome Pathways, Molecular Screening and Toxicogenomics. Available at http://www.oecd.org/env/ehs/testing/molecularscreeningandtoxicogenomics.htm (last accessed June 20, 2014).
29 Organisation for Economic Co-operation and Development (2012) The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No. 168.
30 Vinken, M. (2013) The adverse outcome pathway concept: a pragmatic tool in toxicology. *Regul. Toxicol. Pharmacol.*, **312**, 158–165.
31 Nuclear Regulatory Commission (NRC). Available at http://www.nrc.gov/reading-rm/basic-ref/glossary/risk.html (last accessed June 20, 2014).
32 Food and Drug Administration (2005) Guidance for Industry: Estimating the Maximum Safe Starting Dose in Initial Clinical Trials for Therapeutics in Adult Healthy Volunteers. U.S. Department of Health and Human Services, Center for Drug Evaluation and Research (CDER).
33 National Toxicology Program. Available at http://ntp.niehs.nih.gov/?objectid=72016E39-BDB7-CEBA-FF0CB5AFEA7887A2 (last accessed June 20, 2014).
34 Judson, P.N., Cooke, P.A., Doerrer, N.G., Greene, N., Hanzlik, R.P., Hardy, C., Hartmann, A., Hinchliffe, D., Holder, J., Müller, L., Steger-Hartmann, T., Rothfuss, A., Smith, M., Thomas, K., Vessey, J.D., and Zeiger, E. (2005) Towards the creation of an international toxicology information centre. *Toxicology*, **213**, 117–128.
35 The Carcinogenic Potency Database. Available at http://toxnet.nlm.nih.gov/cpdb/ (last accessed June 20, 2014).
36 International Uniform Chemical Information Database. Available

at http://iuclid.eu/index.php?fuseaction=home.iuclidHome (last accessed June 20, 2014).

37 Bitsch, A., Jacobi, S., Melber, C., Wahnschaffe, U., Simetska, N., and Mangelsdorf, I. (2006) REPDOSE: a database on repeated dose toxicity studies of commercial chemicals – a multifunctional tool. *Regul. Toxicol. Pharmacol.*, **46**, 202–210.

38 Briggs, K., Cases, M., Heard, D.J., Pastor, M., Pognan, F., Sanz, F., Schwab, C.F., Steger-Hartmann, T., Sutter, A., Watson, D.K., and Wichard, J.D. (2012) Inroads to predict *in vivo* toxicology – an introduction to the eTOX project. *Int. J. Mol.*, **13**, 3820–3846.

39 Clinical Data Interchange Standards Consortium. Available at www.cdisc.org/ (last accessed June 20, 2014).

40 Hayashi, M. and Sakuratani, Y. (2012) Development of an evaluation support system for estimating repeated-dose toxicity of chemicals based on chemical structure, in *New Horizons in Predictive Toxicology: Current Status and Application* (ed. A.G.E. Wilson), RSC Publishing, Cambridge, pp. 26–37.

41 World Health Organization (2010) Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment. IPCS Harmonization Project Document No. 9.

42 Niklas, J., Diaz Ochoa, J.G., Bucher, J., and Mauch, K. (2013) Quantitative evaluation and prediction of drug effects and toxicological risk using mechanistic multi-scale models. *Mol. Inform.*, **32**, 14–23.

43 OECD (2007) Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models. OECD Environment Health and Safety Publications, Series on Testing and Assessment, No. 69 (ENV/JM/MONO (2007)2).

44 Hewitt, M. and Ellison, C.M. (2010) Developing the applicability domain of *in silico* models: relevance, importance and methods, in *In Silico Toxicology: Principles and Applications* (eds M.T.D. Cronin and J.C. Madden), RSC Publishing, Cambridge, pp. 301–333.

3
In Silico Approaches: Data Management – Bioinformatics

Arnd Brandenburg, Hans Gmuender, and Timo Wittenberger

3.1
Introduction

With the rapid advances of various high-throughput technologies, the generation of omics data has been established as standard methodology in almost every biomedical field. High-throughput technologies are routinely used for research in basic science, in efforts to understand and treat human diseases, as well as in understanding and prediction of toxic effects of drugs. Omics technologies such as toxicogenomics, which deals with the effects of compounds on gene expression patterns in target cells or tissues, are emerging as key approaches for understanding the mode of action (MOA) of already existing drugs and in screening new drug candidates. Applying omics technologies may reveal, for example, genetic or protein signatures that can be used to predict the short- and long-term biological effects of the exposure to a drug and to identify potential toxic mechanisms with small amounts of compound material at an early stage during drug discovery. Such assessments allow characterizing the mode of action of known compounds and more rapid stop–go decisions during early development stages of new drugs. In addition, toxicogenomics will complement traditional pathology and toxicity studies especially when the clinical outcome of a drug takes a long time to be manifested. Another key step of toxicogenomics is the setup of reference compendia with compounds from different toxicological classes, to classify drug candidates into such reference compendia and to predict their toxicity.

However, as a consequence the amount of biological data is exploding, both in size and in complexity, and large data sets are being generated that have to be stored, statistically analyzed, and interpreted with regard to systems toxicology, not only for each single omics technology but also involving different omics technologies. In addition, such omics data should be correlated with conventional data in order to complete the overall picture of toxic effects. Therefore, increasingly sophisticated computational techniques, efficient means for storing, searching and retrieving data, and powerful algorithms and statistical tools are

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

required. In this chapter, some aspects of the complexity of such systems toxicology approaches will be discussed.

3.2 Experimental Setup and Statistical Power

Before starting experiments, some critical points concerning the experimental setup should be addressed. A major challenge for the selection of the adequate omics technology is to be able to accurately detect analytes present at very different concentrations. In biological samples, transcripts, proteins, and metabolites can be present at concentrations ranging across several orders of magnitude. Repeated experiments help to increase the reliability simply because an observed effect can be confirmed with other biological replicas. More specifically, the statistical power, that is, the chance to reject the null hypothesis (there is no statistically significant difference between sets of given observations) if it is actually false, generally increases with the number of replicas. The number of replicas needed in a given study depends on the size of the effects to be detected and the desired significance level and power of the test. Technical replicas, on the other hand, allow testing the reproducibility of the chosen platform, but nowadays many techniques are usually highly developed and reproducibility is often no longer an issue. However, technical replicas may increase the reliability allowing to average the outcomes and to test for platform-inherent variances. Because the usage of high numbers of biological and/or technical replicas may be limited for economical reasons, great attention should be given to the quality of the samples. Experiments of poor quality detected at any step of the sample preparation process should be eliminated or at least identified. In this context, it is very critical to apply standardized quality assessment protocols and to define quality parameters and thresholds beforehand.

A critical point for assessing the potential toxicity of compounds is a suitable study design with the inclusion of meaningful time points and compound concentrations at which changes of, for example, gene expression are expected. For an adequate selection of concentrations, it has to be considered whether primary, causal, and subtle responses to the drugs should be measured, or secondary responses that are not necessarily compound specific but rather more general reactions of the system to perturbances (e.g., adverse outcome pathways). Determining such optimal parameters and reducing the effects of distorting environmental influences and sources of technical artifacts will help to optimize the quality, reproducibility, and power of the studies. Finally, one should not forget to include suitable positive and negative controls (e.g., stress, vehicles). However, omics technologies and protocols have changed and will continue to evolve over time, hence presenting the challenge of combining and analyzing data from different development stages of a given platform. For example, RNA isolation and hybridization protocols as well as the design of microarrays have changed and were improved over

time. Such changes should be taken into account when comparing results from different periods.

3.3 Properties of Different Omics Data

3.3.1 Next-Generation Sequencing Data

Deep sequencing is now frequently used to quantify total mRNA abundance levels, as its wide dynamic range and high signal-to-noise ratio facilitate the quantification of low-abundance transcripts [1]. Whole-transcriptome analysis using next-generation sequencing (NGS) technologies, that is, RNA-seq, represents a new possibility to analyze gene expression. In several gene expression studies, microarrays are now being replaced by RNA sequencing-based methods, which can identify and quantify rare transcripts without prior knowledge of a particular gene and can provide information regarding alternative splicing and sequence variations in identified genes. The variety of NGS technologies makes it likely that multiple platforms will coexist in the future, with some having clear advantages for particular applications over others. A challenge will be to integrate and analyze NGS and microarray data together. For microarrays, the abundance of transcripts is measured as fluorescence intensities, which is a continuous response, whereas for next-generation sequencing data the abundance is measured as count data (number of reads). Therefore, procedures that are successful for microarray data analysis are not directly applicable to NGS data. Similar to the MicroArray Quality Control (MAQC-III) consortium effort [2], projects comparing NGS and microarray data (e.g., SEquencing Quality Control (SEQC)) are underway. The goal is to determine whether the two platforms have a similar ability to detect differences between various perturbed biological systems. The questions are whether both technologies perform unequivocally or whether one outperforms the other in a certain range. Another question, specifically interesting for systems toxicology, is whether the cross-platform concordance is highly dependent on the transcription response caused by the compound. Is the overlap of differentially expressed genes between the two platforms larger in a more disturbed system? Has NGS less difficulties to accurately quantify low expression?

Different data types imply also different statistical characteristics of the data sets. While microarray data are assumed to follow a normal distribution, NGS data are count data and hence not normally distributed but follow a Poisson or negative binomial distribution. Therefore, log transformations usually applied to gene expression intensities derived from microarrays should not be carried out with NGS data, in particular because the information about zero counts would be lost. Appropriate statistical methods that implement the correct assumptions about the data distributions should be used. Popular statistical tests to look for

differential expression of RNA-seq data include the edgeR [3] and DESeq [4] algorithms. In addition, specific normalization methods have to be applied, for example, to correct for different library sizes (number of sequenced fragments).

3.3.2 DNA Methylation Data

Another application of NGS is genome-wide DNA methylation profiling. It is well established that epigenetic hyper- or hypomethylation of, for example, promoter regions may regulate gene expression. Deep sequencing of bisulfite-treated DNA allows for the quantification of changes in the methylation status of CpG islands at base-pair resolution [5,6]. Bisulfite sequencing examines cytosine DNA methylations at nucleotide resolution along single DNA strands, yielding single-nucleotide resolution information about the methylation status of a segment of DNA. Treatment of DNA with bisulfite converts cytosine residues to uracil, but leaves 5-methylcytosine (and 5-hydroxymethylcytosine) residues unaffected. The distribution of methylation levels is usually expressed as β-values (percentages of methylation) or M-values ($\log_2$ ratios of the intensities of methylated versus unmethylated probes). The β-value has a more intuitive biological interpretation (% methylation) but shows severe heteroscedasticity; therefore, it is proposed that M-values are statistically more valid. The relationship between the β-value and the M-value is a logit transformation [7]. DNA is mainly either fully hypermethylated or fully hypomethylated, but seldom in-between; therefore, methylation data do not follow a normal distribution.

3.3.3 miRNA Data

Nowadays it is well established that biological systems use a variety of mechanisms to maintain their functions after environmental and genetic perturbations. Increasing evidence suggests that microRNAs (miRNAs) interfere with biological processes by controlling posttranscriptional gene expression, reinforcing transcriptional programs, and attenuating aberrant transcripts, and they may help suppress random fluctuations in transcript copy number [8]. Specific microarrays as well as sequencing allow the measures of miRNAs and such measurements may also help to understand and interpret biological processes.

3.3.4 CNV and SNP Data

Low-depth quantitative DNA sequencing enables a rapid analysis of copy number variations (CNVs), both at whole genome level and at gene level [9,10], as well as the detection of chromosomal rearrangements. Alignment of DNA sequence reads to a reference genome further allows for identification of single-nucleotide polymorphisms (SNPs), InDels, and mutations, of both germline

(with only low read depth needed) and somatic origin (high read depth required). Characterization of SNPs is usually a two-step process, involving "SNP calling" to identify variable sites and "genotype calling" to determine the genotype [11,12].

3.3.5 ChIP-seq Data

ChIP-seq, which combines chromatin immunoprecipitation (ChIP) with NGS, will probably replace ChIP-chip technologies and allows interrogating whole-genome protein–DNA interactions as, for example, histone–DNA or transcription factor–DNA interactions. In a ChIP-seq experiment, the DNA fragments from binding sites of a target protein are enriched through immunoprecipitation. Sequenced reads of such immunoprecipitated DNA fragments are aligned to a reference genome to identify sites of enrichment. A special case of ChIP-seq is MeDIP-seq, by which methylated DNA is precipitated and sequenced.

3.3.6 Gene Expression Microarray Data (Affymetrix)

In this section, we will mainly discuss Affymetrix gene expression microarrays; other vendors provide their own solution for the transformation of fluorescence intensities into gene expression values.

Affymetrix GeneChip technology measures gene expression using hybridization of cRNA to 25-mer oligonucleotides. Typically, a mRNA molecule of interest (usually related to a gene) is represented by a probe set composed of 11–20 probe pairs of these oligonucleotides. Each probe pair is composed of a perfect match (PM) probe, a section of the mRNA molecule of interest, and a mismatch (MM) probe that is created by changing the 13th base of the PM with the intention of measuring nonspecific binding.

Summarizing the probe intensities of probe sets is a critical preprocessing step for the expression analysis based on Affymetrix GeneChip technology. A great variety of condensing algorithms based on different approaches exist. A popular representative is the statistical condensing algorithm (MAS5, microarray suite) provided by Affymetrix in its analysis software [13]. It is a single-array condensing algorithm using the PM and MM intensities for each probe on the arrays. An advantage is that it can be applied independent of the data sets. Li–Wong condensing is an alternative that condenses simultaneously sets of chips [14]. The key idea of this condensing is that probe weights are learned from the input data and a weighted average of the probe sets is computed. RMA (robust multiarray analysis) condensing is also inspired by this idea, but represents an improvement in that the model is designed to be more robust against outliers. The authors of the RMA model have demonstrated using an example data set that RMA condensing is superior to Affymetrix statistical and Li–Wong condensing [15–17]. They conclude that a condensing algorithm should only rely on the PM

intensities and not use the MM intensities, since the latter have frequently a large contribution due to hybridizations to target genes. Furthermore, they find that the contributions of the different PM probes should be weighted in order to cope with the different hybridization affinities and cross-hybridizations of the PM probes within a probe set. RMA condensing learns these probe weights from the input data and the expression index is never negative. They achieve this by using a Bayesian model for the hybridization intensity. The GC-RMA condensing is a modification of the RMA condensing that uses the GC content of probes to determine the background. A disadvantage of the simultaneous condensing algorithms is that new samples cannot be added seamlessly to an existing data set, but the new data set has to be condensed again as a whole.

In addition to the well-established expression analysis microarrays, Affymetrix and others also provide, for example, whole-transcript expression arrays, tiling arrays, exon arrays, and so on that need in general different or slightly different condensing and analysis algorithms.

A further point to be considered is the annotations of the probes or probe sets. The annotations will change over time gaining new insights into the genomic information of a species. Therefore, it cannot be avoided that microarray data should be recondensed from time to time because the changing combination of probes to probe sets and the varying annotations may lead to different gene expression level results.

Standard interpretations of statistical tests such as the Student's *t*-test assume that the data are sampled from normal populations with equal variances. Expression data from microarrays are usually not normally distributed, but a logarithmic transformation can lead to an approximate normal distribution and is therefore a recommended procedure before performing statistical tests [18].

3.3.7 Mass Spectrometry Data

Mass spectrometry has become an important technology to obtain high-quality data on metabolites and proteins in body fluids or tissues. For metabolomics, LC–MS and GC–MS are the primary technologies. For proteomics, LC–MS/MS is widely used for a range of applications. The principles of the main processing steps are quite similar for the different data types and technologies and are also largely independent of the specific machine that produces the data, but clearly the most suitable algorithms and parameters have to be adapted to the case at hand. The processing of the raw profile data should start with the removal of noise and background signals. To reduce the noise, three strategies can be applied: data smoothing, chemical noise subtraction, and small structure removal.

Data smoothing helps to remove possible irregular peak profiles across scans. If chromatograms have been generated from blank samples, they can be subtracted from the chromatograms of the biological samples. Electronic noise, manifesting itself by small structures in m/z and RT coordinates, can easily be identified and removed. Chemical noise can be generated, for example, by

column bleeding and is identified by the increased background level in RT direction at several m/z values. To remove this RT background, an estimation of its level can be performed by computing a specified quantile value in a window centered at the subtraction position. The remaining noise can additionally be filtered by removing a constant intensity value. An adequate noise subtraction step is critical for correctly removing the noise while preserving the relevant signals and thus allowing their correct identification.

Due to nonexact reproducibility of the chromatography across experiments, the cleaned data need to be aligned so that peaks generated by the same compound are shown at the same position in different chromatograms. A correction for retention time shifts between chromatograms can be achieved by applying a (nonlinear) transformation to each individual retention time, mapping the original time onto a common universal retention time. For LC–MS and GC–MS metabolomic experiments, the resulting RT corrections are often quite small, while for proteomic data sets the RT corrections can be large. Shifts in m/z values across chromatograms can be corrected by using lock masses, if available.

Detection of peaks can be performed on the result of averaging all data points across samples. The detected peaks should then be grouped into isotope clusters, for LC–MS metabolomic data best by using a library of compounds as a reference for the possible isotopic envelopes. The library can be generated, for example, from the HMDB database (the Human Metabolome Database, www.hmdb.ca/), which can also serve as the source for the compound identification step. If fragment MS/MS spectra are available or in the case of GC–MS data, spectral libraries of known compounds can be used to identify metabolites by comparing the measured spectra with the entries in the libraries [19]. A number of such libraries are available from a variety of sources: the NIST Library for GC spectra [20], the Fiehn Library [21], and the Golm Metabolome Database [22] are some popular examples. Each measured spectrum is compared against each library spectrum and for each of the comparisons a score is computed that quantifies the similarity of the spectra.

One notable difference between metabolite data and proteomic data is the occurrence of highly charged peptides in the latter, leading to more complex isotope cluster patterns of definite charge. If available, theoretical peptide isotope intensity profiles can be used to improve the clustering.

The secondary MS/MS scans form the basis of the identification of the peptides and proteins. The fragment scans can be compared with a library of peptide fragment spectra. Popular library search engines offering customizable options for choosing the relevant organism, the methods for matching of the spectra, acceptance thresholds, and so on include Mascot (Mascot search engine, www.matrixscience.com/), Sequest (Sequest homepage, http://fields.scripps.edu/sequest/), and X!Tandem (X! search engine, http://www.thegpm.org/tandem/). Posttranslational modifications of proteins can be detected in this way as well.

The absolute abundance of an isotope cluster can be estimated by taking the maximal signal intensity of all peaks in the cluster, by summing up all measured

intensities within the boundaries of all contributing peaks, or by calculating the sum of the volumes of all contributing peaks.

If multiple clusters are annotated with the same compound name, possibly due to peak splitting but potentially also due to mismatches, the corresponding intensities should be summed together. To quantify the abundance of a peptide, the different charge states should be grouped together and the intensities should be summed up.

Finally, a logarithmic transformation should be applied before further data analysis steps are taken, since MS data are approximately log-normally distributed.

3.3.8 Missing Values and Zero Values

It is common that some probes or spots on microarrays cannot be measured resulting in missing values. The number of missing values can even increase if different filters are applied to the data, such as filtering out saturated spots, probe sets below the so-called detection *p*-value threshold from MAS5, spots with very high background, and so on. For data acquired by mass spectrometry, missing values can result from the logarithmic transformation of zero values generated by noise reduction steps. In some cases, it is useful to replace such missing values by imputing. In the row mean imputation, for each row of the data matrix an average value (usually the geometric mean or the median) is calculated from the non-missing values and all missing values in the row are replaced by this average. Many more sophisticated imputation methods have been invented [23]. As briefly discussed earlier, zero values from count data should not be treated as missing values since they contain important information.

3.3.9 Data Normalization

There are many reasons why data have to be normalized, including, for example, unequal quantities of starting RNA, differences in labeling or detection efficiencies, batch effects, and systematic biases in the measured expression levels. Typically, the first transformation applied to expression data adjusts the individual hybridization intensities to balance them appropriately so that meaningful biological comparisons can be made. There are many approaches possible to normalize data. One of the simplest assumes that approximately the same number of labeled molecules from each sample should hybridize to the arrays and the total hybridization intensities summed over all elements in the arrays should be the same for each sample. A central intensity normalization scales the individual intensities so that the geometric mean or median intensities are the same across all arrays. Another common normalization procedure transforms absolute intensities into relative data by dividing, for example, the intensities of a group of samples after a treatment with the average of a time-matched control group

(relative normalization). More sophisticated normalization procedures are LOWESS normalization, *Z*-transformation, quantile normalization, median polish, and so on. The LOWESS (locally weighted regression and smoothing scatter plots) normalization applies a nonlinear correction to the selected data matrix. A local polynomial regression model is fitted to each point and the points close to it. The method is also sometimes referred to as LOWESS smoothing. The smoothed data usually provide a clearer picture of the overall shape of the relationship between the x and y variables [24]. The *Z*-transformation transforms the log expression profile of each gene in such a way that it follows a normal distribution with zero mean and a standard deviation equal to 1. Quantile normalization is a technique for making two distributions identical in their statistical properties. As normalization it can often be used to remove small differences between measured distributions when there is a common underlying distribution (based on theoretical arguments or based on analysis of the data). Taken to the extreme, it can also transform any given data vector such that its data follow a given model distribution [15]. The median polish algorithm [25] iteratively subtracts the row and column medians of the data matrix until the residual values stabilize. In general, normalization is a critical step in the analysis of omics data. It should, however, always be checked carefully if the underlying assumptions of a given normalization method are fulfilled by the data at hand. For example, subsets of genes selected based on some biological criteria may violate the assumption of a common median expression across all samples.

3.4 Statistical Methods

3.4.1 Data Overviews

There are some useful analyses to get an overview of data sets, to detect outliers, and to detect structures in the data. Such structures or clusters can be due to artifacts such as batch effects, fluorescent dye effects, and so on, but also due to factors relevant for further analyses, such as sex, chemical measures, phenotypic anchoring such as histopathology, drug concentration and/or time effects, and so on. Such random or fixed factors can be taken into consideration by performing, for example, linear models or for a stratification of the data set. Analyses providing data overviews are, for example, histograms displaying the distribution of expression values, box plots displaying the distribution of the values for each experiment, bar charts, clustering methods such as hierarchical clustering (and using heat maps for visualization), and so on. For clustering methods, there are different possibilities to measure distances, for example, correlation, Euclidean, normalized Euclidean, and Manhattan, and to determine the linkages, such as complete, average, single, and so on. Quite often, principal component analyses (PCA) are very informative. PCA is a classical means of unsupervised

(i.e., non-hypothesis driven, in contrast to, for example, partial least square analysis) dimensionality reduction and visualization of multivariate data. It involves a mathematical procedure that transforms the covariance matrix into a diagonal form. The new coordinates are called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Restriction to the first two or three principal components often is sufficient for capturing dominant structures of the data and uncovering important relationships between individual experiments or groups of experiments.

3.4.2 Null Hypothesis/Type I and Type II Errors

In omics experiments, thousands of features (genes, proteins, metabolites, etc.) are measured simultaneously across different experimental conditions. Two important issues arise in identifying differentially expressed items: Which items are differentially expressed? What is the quantitative statistical evidence for the possible differential expression of the items?

The standard tools to answer these questions are as follows: rank the items by computing *p*-values using a suitable test statistic and use multiple testing methods to assess the error rates in a given list of items (false positives and false negatives).

The *p*-value of a statistical test is the probability that random data yield a value of the test statistic that is equal to or more extreme than the observed value. For example, in the case of a *t*-test, the (two-sided) *p*-value is the probability of getting a value for the *t*-statistic from random data that has an absolute value equal to or larger than the observed one. In this way, the null hypothesis H_0, which states that there is no effect in the data, is tested. For the *t*-test, the null hypothesis states that the mean values of the two groups are equal. The type I error or false-positive rate (often denoted by α) corresponds to the probability that H_0 is rejected although it is true. The type II error or false-negative rate (denoted by β) corresponds to the probability that H_0 is accepted although it is false. In the context of microarray data, a large false-positive rate means that many truly null genes are classified as being differentially expressed, while a large false-negative rate means that many interesting differentially expressed genes are not identified. The power or sensitivity of a test is given by $1-\beta$ and corresponds to the ability of the test to reject the null hypothesis (to find interesting genes) when it is actually false (when these genes are actually differentially expressed).

3.4.3 Multiple Testing Methods

The larger the number of items tested for differential expression, the more important it becomes to apply multiple testing methods. In the special context of microarray analyses and more general in all cases where the number of

unchanged items is large compared with the number of changed items, it turns out that the so-called false discovery rate (FDR) is an appropriate error rate. It is defined as the ratio of the number of genes that are called significant although the null hypothesis is true and the total number of genes that are called significant. Both quantities depend on the *p*-value threshold one sets. The numerator can be estimated in different ways. The most common method is due to Benjamini and Hochberg [26]. The estimate of Storey and Tibshirani [27] is a modification that constitutes a less conservative estimate of the FDR. As a rule of thumb, one can keep in mind that for a total number of m genes and a p-value threshold Θ, the number of genes from a random distribution that have a p-value below Θ is equal to $m\Theta$. If the actual number of genes with p-value below the chosen significant threshold is not well above that number, the FDR will be very high for that group of genes.

In addition to the previously described methods, an estimate of the FDR based on permutations (balanced or not balanced) of the experiments can be made.

3.4.4 Statistical Tests

If it can be assumed that the data are normally distributed with equal variances, a *t*-test or an ANOVA (analysis of variance) is the most commonly used method to evaluate the differences in means between two or more groups, respectively. The *p*-level represents the probability that the observed – or a more extreme – value of the test statistic is due to chance, provided the null hypothesis is true. As described earlier, the larger the number of genes tested for differential expression, the more important it becomes to apply multiple testing methods. The Welch test generalizes the *t*-test by dropping the assumption that the variance is identical in both groups. Nonparametric tests use ranks rather than the actual measured values. The Wilcoxon test (also known as the Mann–Whitney test) works for two groups and computes as test statistics partial rank sums. This type of test makes no assumptions on the distribution of values within the groups. The Kruskal–Wallis test represents an extension of the Wilcoxon test to more than two groups. Finally, the Kolmogorov–Smirnov test tests whether the distribution of values in the two groups differs. It makes no assumption on the distributions itself and is therefore very general. However, the disadvantage is that this test requires more input data than, for example, the *t*-test.

As mentioned earlier, NGS data are not normally distributed and therefore more adequate tests should be applied to analyze such data.

3.4.5 Linear Models and Linear Mixed Models

The linear model analysis allows to describe data by a linear combination of experimental factors and to study the impact of each of these factors on the quality of the model. These can be so-called fixed factors, the levels of which the

scientist "fixed" when designing the experiment in order to learn about their influence on the data, random factors that are often uncontrollable and also called "nuisance" factors, or covariates that are concomitant variables that contribute to the variation in the data. The effects of the factors and their interactions are estimated and *p*-values are computed for the null hypothesis that states that, roughly speaking, a certain factor has no influence on the data. The linear model analysis is a very flexible and powerful statistical method that includes as special cases the ANOVA for an arbitrary number of factors, with or without interactions, the analysis of covariance (ANCOVA), and linear regression. Batch effects can be modeled as random factors in a linear mixed model. In this way, their influence on the interesting fixed factors can be eliminated. An advantage over batch removal via explicit data transformation methods is that in linear mixed models the degrees of freedom are correctly taken into account in the computation of the *p*-value. Paired designs are also covered by linear models, with the pairing variable (e.g., patient) simply treated as a random factor.

Further generalizations of linear models are the so-called generalized linear models. They allow modeling the data by nonlinear relationships between the estimated mean and the covariates and also with non-Gaussian fluctuations around the mean. Important applications are statistical tests for count data obtained in RNA-seq experiments. Here, a so-called log-linear model is used that guarantees that the estimated mean is positive, together with the assumption that the fluctuations around the mean follow a negative binomial distribution. In addition, so-called offsets can be included that account for different library sizes as well as an additional scaling normalization [3,4,28].

3.5 Prediction and Classification

3.5.1 Overview

One approach in systems toxicology is to build a database of expression signatures for known compounds. As the goal is to profile novel compounds against this reference compendium, it is important to have several members of each toxicological class represented. There is no limitation to the number of compounds that can be profiled, and as long as the experimental procedure and the type of microarrays used are not significantly changed, one can continuously add more compounds to the database to improve and enlarge its predictive quality. Under certain circumstances, it is even possible to include not only gene expression, but also proteomic data as long as values of the data sets are comparable. Not only gene expression patterns but also additional information from blood samples, phenotypic observations, tissue sections, clinical outcomes, and so on should be stored because this information may be very useful to build classes for classification algorithms and for the interpretation of gene expression patterns.

A basic distinction for constructing a reference compendium can be made between methods that take into account all the available expression data and methods that use only a limited number of previously determined marker genes. The advantage of using a method with no restriction on the number of genes is that more genes and pathways can be utilized for toxicological predictions as the list of compounds extends. Therefore, to get as much information as possible, such "unlimited" approaches should be preferred whenever possible. However, not every laboratory will have the resources for large arrays with thousands of genes, and some toxicology laboratories may turn to smaller, customized chips or qRT-PCR on subsets of genes for large-scale studies. A compromise may be to use large arrays to determine an optimal set of genes that allow to classify a large variety of compounds, and to switch later to a smaller, customized array with that set of genes. Marker gene selection should be based on unbiased computational methods employing algorithms to select genes with high discriminatory power between different compound classes. The advantage of using such a method is that even uncharacterized pathways are taken into consideration. However, previously acquired knowledge on the effects of compounds on cellular pathways can be used to enrich the set of statistically determined marker genes.

Once a reference compendium has been established and cross-validated, it can be applied to classify novel compounds. Matching a gene expression profile of a new drug candidate to profiles of known compounds can then reveal whether it will most probably show undesired side effects. Novel compounds that have been unequivocally classified and properly annotated can then be added to the reference compendium, provided that the quality of the experiments is at the same high level as for those used to construct the reference compendium. The cycle is repeated until the compendium achieves a robustness that allows it to reliably classify novel compounds or even compound mixtures.

3.5.2 Generating a Reference Compendium of Compounds

For a given toxicological endpoint, such as liver necrosis, replicate experiments are performed for a multitude of compound dosages and treatment times. Two basic approaches exist for structuring such large and complex data sets as they occur in toxicogenomic experiments – unsupervised learning methods and supervised learning methods.

In unsupervised learning methods, no assumption is made about the structure of the data. Two-dimensional hierarchical clustering is an example of unsupervised learning. If genes that distinguish between toxic MOAs have been carefully selected, the results of such analyses can be displayed as "toxic fingerprints." However, this approach may lead to unsatisfactory results if effects due to the experimental setup (technical or biological influences that result in experimental noise) override compound-specific effects.

The second approach, supervised learning, takes into account the available information from well-established compounds. This approach has been shown

to structure toxicogenomic data meaningfully. Automated prediction of toxicity relies on classification algorithms that use the expression data obtained from treated tissues to make reliable predictions. Furthermore, gene selection algorithms allow researchers to focus on genes that are indicative of the toxic effects of interest, thereby optimizing the predictive strength.

Classification can follow two different strategies. The first strategy uses as many genes as possible – preferably all available genes on a microarray. A classification algorithm is applied that works well in high-dimensional spaces, is robust, and remains insensitive to experimental noise. When processing such a high-dimensional gene space, supervised learning algorithms select planes that optimally separate the compound classes. A prominent example of such a classifier is the support vector machine algorithm [29]. Generalized forms of linear discriminant analysis, *K*-nearest neighbors, and decision trees can also be used as classifiers. The support vector machine algorithm is based on constructing separating hyperplanes with a maximal margin. In principle, misclassifications of training set elements are allowed, which is controlled by a penalty parameter. It determines the amount with which the misclassification of a training set element is penalized. The second strategy employs an emphasis on careful gene selection. Genes can be selected manually according to biological knowledge or with the help of statistical algorithms. There is no universal answer to the question which of the two strategies is better. Dispensing gene selection altogether is certainly the concept that is easier to extend, because when using a restricted set of marker genes, an update of the reference compendium will probably change the marker gene set significantly. On the other hand, the gene selection approach might give further insight into the nature of the cellular processes involved. A strict decision for the first or second strategy is not recommended. Instead, it is recommended to work with a substantial arsenal of classification algorithms and gene selection methods and to experimentally determine the strategy and the algorithms that best suit the data.

Experience has shown that employing the first strategy using all the genes and then subsequently reducing the number of genes using statistical algorithms until an optimal gene set is found is a promising approach. For gene selection, a standardized algorithmic approach is less biased than a subjective gene selection based on established knowledge of marker genes. Gene selection algorithms range from simple univariate tests, such as one-way analysis of variance, to highly sophisticated ones, for example, recursive feature elimination [30].

3.5.3 Cross-Validation

Cross-validation is an important technique for deciding which classification algorithms are optimal. Furthermore, cross-validation can help to discover unexpected relationships between MOAs. A classifier separates, based on training data, the gene expression space into regions, the domains of the different groups. In a cross-validation step, the data of one or several experiments are taken out of

the training data set and the classifier is recomputed. If the experiment that was excluded is reassigned to its original MOA class, the assignment is considered "stable under cross-validation." Another experiment, after removing and recomputing, may be assigned to a different class indicating a "cross-validation error." Several methods of cross-validation are available. The "leave-one-out" cross-validation method removes one experiment at a time. With "*n*-fold" cross-validation, a randomly selected fraction of experiments is removed and reclassified. For toxicological experiments, "subgroup" cross-validation can be employed when further categorizing and classifying experiments. An example for "subgroup" classification is to classify in the first round according to histopathological endpoints and in the second round according to compound MOA classes. Especially if replicas are available, removing the replicas as a group results in a more realistic cross-validation, because the probability that an experiment classifies into the correct category may be much higher if only one experiment of the replicas is removed and reclassified. Using cross-validation methods, the experiments that fail to classify correctly are used to calculate the cross-validation error rate. By systematically assessing various classification and gene selection methods, one is then in the position to find the combination of classifier and gene selection method, as well as the size of the gene set, with the optimal predictive power. The reference compendium derived from an optimal set of marker genes and an optimal classification algorithm can then be used to classify novel compounds.

3.5.4 Selection Bias

If the selection of genes is based on the same data as the data on which the misclassification rate is estimated, the problem of a selection bias will be encountered. If one selects, for example, the top-scoring ANOVA genes and then calculates the misclassification rate on the same experiments, the estimate of the misclassification rate will usually be much smaller than the error rate one would get on independent experiments.

3.6 Combining Different Omics Data and Biological Interpretations

The ultimate goal of systems toxicology is obviously to combine different omics technologies into an overall picture of biological processes and to understand in detail the mode of actions as effect of the perturbances of, for example, a drug. However, this is still not a task that can be fully automated. Searching for correlations between methylated DNA regions and gene expression, or gene expression and correlated protein levels, may give some hints but surely not the full picture. It can well be that gene expression is increased but not reflected at the measured protein level. Therefore, the inevitable method is in most cases to generate a working hypothesis based on markers from a given omics technology and

to integrate in usually rather laborious detailed work additional pieces of information obtained from other omics data or other markers (see Refs [31–33]). Possibilities to interpret omics data are, for example, analyzing promoter regions for transcription factor binding sites, searching for hyper- or hypomethylated CpG islands, searching for CNVs, SNPs, and miRNAs with potential regulation capacities, and so on. In this context, it has to be mentioned that standard procedures to validate omics results with other technologies such as PCR, Western blot, and so on are very common and indispensable.

However, many analysis and visualization tools have been developed in parallel to the development of the omics technologies. In recent years, the focus in expression data analysis has shifted from single gene to the gene set level analyses. This change has been motivated by realizing that many diseases are associated with regulations in a set of related genes rather than an increase in a single gene [34]. Gene set analyses are also expected to facilitate the interpretation of lists of differentially expressed genes. Several recent reviews provide overviews of the developed methods for such gene set analysis [35,36]. The methods can be categorized as "self-contained" or "competitive." Self-contained methods analyze the association between the phenotype and expression in the gene set of interest while ignoring genes not in the gene set. Competitive methods compare the gene set with its complement in terms of association with the phenotype. Examples include the popular GSEA algorithm [34], GSA (gene set analysis) [37], SAFE [38], and random set methods [39]. There are many sources providing gene sets such as gene ontology (GO) categories, metabolic pathways, signaling pathways, information about protein–protein interactions, and so on.

A generic method to find associations between groups of differentially expressed genes and gene sets defined by some common property is Fisher's exact test. It yields the probability that observed counts in a so-called contingency table are due to pure chance and thus allows a rejection of the null hypothesis (independence of regulated genes and a given biological category) on the basis of a p-value threshold. It belongs to the class of exact tests, so called because the significance of the deviation from the null hypothesis can be calculated exactly using combinatorics, rather than relying on an assumption about the null distribution of the data.

3.7 Data Management

Often it is not the limited availability of omics data that prevents the scientific community from making optimal use of published studies (or, to mention another example, companies from making optimal use of their legacy data). Rather, it is the description of the data, the experimental design, and the availability of other non-omics data that are difficult to interpret, incomplete, or erroneous. Public efforts to standardize such metadata have greatly facilitated the

reuse of data within the scientific community, but are mainly focused on technologies such as microarrays (MIAME, http://www.mged.org/Workgroups/MIAME/miame.html) or proteomics (PRIDE XML, http://www.ebi.ac.uk/pride/help/archive/submission/pridexml). The tranSMART initiative (http://transmartfoundation.org), on the other hand, develops a platform for the organization of all clinically relevant data, including omics, while diXa (http://www.dixa-fp7.eu) has a similar approach for toxicogenomic data sets, both enabling data analysis across multiple studies (meta-analyses) on a molecular or pathway level, respectively. Finally, the ISA-TAB framework (http://www.isa-tools.org) implements a standard for metadata organization that is applicable across a wide variety of application domains and platforms. It is implemented in EBI's BioInvestigation Index (http://www.ebi.ac.uk/bioinvindex) and currently adopted for more specific domains, such as nanotechnology (https://wiki.nci.nih.gov/display/ICR/ISA-TAB-Nano). All of these approaches rely on the use of standardized vocabulary, and accordingly there are a large variety of domain-specific ontologies.

The generic problem addressed by all metadata approaches is one of data organization: The data analyst needs to understand the experimental design, the (biological and chemical) entities, their function and use in a study, and the factors that – intentionally or confounding – influence the measurement results. The metadata also have to describe the exact relationship between data and study entities (e.g., protocols used on samples derived from study subjects to obtain the data).

The following paragraphs outline the generic use cases and specifications of an infrastructure that supports data organization in that sense.

For users with a data analysis focus, the description of the study design has to include the data relevant for analysis and interpretation. Such a user will primarily use the metadata to document all parameters that are intentionally varied, or controlled, within a given study. Other parameters (such as the concentration of a given buffer) are kept constant and will not be relevant for data analysis in most cases; a reference to a protocol might suffice here. However, if it turns out that another parameter influences the outcome of the experiment, this confounding variable needs to be used in the analysis process as well. Therefore, flexible adaptation of the metadata with new parameters (or treatment schedules) is necessary.

A user interested primarily in data organization might have a much broader interest in data and information associated with a given study. This might include also raw data, technical parameters such as batch IDs and instrument settings, and other (constant) experiment protocol parameters. Users might want to "collect" these data, or they might want to "drill down" these data on request.

Some core functionalities of a system supporting these use cases are as follows:

- Definition/creation of study design layouts (using data types below).
- Organization of data:
  - Assignment of data sets to study objects.

  - Assignment of annotation data to study objects.
- Management of study design and annotation data:
  - Storage of data.
  - Import from external sources.
  - Validity and consistency checks, including controlled vocabulary/ontologies.
- Data browsing and searching (including study design visualization).
- Data export and download:
  - Study design data.
  - Annotation data.
  - Measurement data.
- Support for structuring data sets:
  - Loading of annotation (also inherited from higher-order entities).
  - Grouping of data sets.
  - Selection of factors for analysis.

Apart from general study design information, that is, treatment, sampling, and measurement protocols, the system needs specifically to keep track of study design parameters and factors relevant for the data analysis. Further, all non-omics measurements, including patient data, have to be managed in a way that they can be used either as (numerical or categorical) annotation during analysis of omics data (e.g., as phenotypic anchors) or as data matrices to be analyzed in parallel to the omics data.

Such a data management system consists of several components:

- A database to manage the metadata and the access to measurement data (but not necessarily the measurement data themselves, since these are often stored in pre-existing repositories).
- A "study design wizard" that facilitates the reporting of study information and metadata.
- Data upload/submission/importing tools.
- Data searching/browsing/exporting tools.
- File format converters.
- APIs to link to external data storages and ontologies.

The effort needed by data submitters to report metadata completely and structured is probably the biggest hurdle to overcome. However, these efforts are still low compared with the curation efforts needed on unstructured or even incompletely reported metadata, if reuse by others can be achieved at all.

The ISA-TAB tools collection (http://isatab.sourceforge.net/tools.html) encompasses most of the components above, including "IsaCreator" helping data submitters to report their metadata well organized and standardized.

References

1 Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.

2 Shi, L., Campbell, G., Jones, W.D., Campagne, F., Wen, Z., Walker, S.J. *et al.* (MAQC Consortium) (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.*, **28**, 827–838.

3 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

4 Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.

5 Bock, C. (2012) Analysing and interpreting DNA methylation data. *Nat. Rev. Genet.*, **13**, 705–719.

6 Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G., and Meissner, A. (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.*, **28**, 1106–1114.

7 Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010) Comparison of Beta-value and *M*-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, **11**, 587.

8 Ebert, M.S. and Sharp, P.A. (2012) Roles for microRNAs in conferring robustness to biological processes. *Cell*, **149**, 515–524.

9 Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E. (2012) Copy number variation detection and genotyping from exome sequence data. *Genome Res.*, **22**, 1525–1532.

10 Teo, S.M., Pawitan, Y., Ku, C.S., Chia, K.S., and Salim, A. (2012) Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*, **28**, 2711–2718.

11 Nielsen, R., Paul, J.S., Albrechtsen, A., and Song, Y.S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.*, **12**, 443–451.

12 Schadt, E.E., Woo, S., and Hao, K. (2012) Bayesian method to predict individual SNP genotypes from gene expression data. *Nat. Genet.*, **44**, 603–609.

13 Hubbell, E., Liu, W.M., and Mei, R. (2002) Robust estimators for expression analysis. *Bioinformatics*, **18**, 1585–1592.

14 Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA*, **98**, 31–36.

15 Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

16 Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003a) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.*, **31**, e15.

17 Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003b) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

18 Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.*, **32** (Suppl. 2), 496–501.

19 Stein, S.E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, **10**, 770–781.

20 Babushok, V.I., Linstrom, P.J., Reed, J.J., Zenkevich, I.G., Brown, R.L., Mallard, W.G.,and Stein, S.E. (2007) Development of a database of gas chromatographic retention properties of organic

compounds. *J. Chromatogr. A*, **1157**, 414–421.
21 Kind, T., Wohlgemuth, G., Lee, D.Y., Lu, Y., Palazoglu, M., Shahbaz, S., and Fiehn, O. (2009) FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal. Chem.*, **81**, 10038–10048.
22 Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmüller, E., Dörmann, P., Weckwerth, W., Gibon, Y., Stitt, M. *et al.* (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **21**, 1635–1638.
23 Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R.B. (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics*, **17**, 520–525.
24 Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.*, **74**, 829.
25 Hoaglin, D.C. and Mosteller, F. (1983) *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, Inc., New York.
26 Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.
27 Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA*, **100**, 9440–9445.
28 Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.
29 Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines*, Cambridge University Press.
30 Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002) Gene selection for cancer classification using support vector machines. *Mach. Learn.*, **46**, 389–422.
31 Boitier, E., Amberg, A., Barbié, V., Blichenberg, A., Brandenburg, A., Gmuender, H., Gruhler, A., McCarthy, D., Meyer, K., Riefke, B. *et al.* (2011) A comparative integrated transcript analysis and functional characterization of differential mechanisms for induction of liver hypertrophy in the rat. *Toxicol. Appl. Pharmacol.*, **252**, 85–96.
32 Ellinger-Ziegelbauer, H., Adler, M., Amberg, A., Brandenburg, A., Callanan, J.J., Connor, S., Fountoulakis, M., Gmuender, H., Gruhler, A., Hewitt, P. *et al.* (2011) The enhanced value of combining conventional and "omics" analyses in early assessment of drug-induced hepatobiliary injury. *Toxicol. Appl. Pharmacol.*, **252**, 97–111.
33 Matheis, K.A., Com, E., Gautier, J.-C., Guerreiro, N., Brandenburg, A., Gmuender, H., Sposny, A., Hewitt, P., Amberg, A., Boernsen, O. *et al.* (2011) Cross-study and cross-omics comparisons of three nephrotoxic compounds reveal mechanistic insights and new candidate biomarkers. *Toxicol. Appl. Pharmacol.*, **252**, 112–122.
34 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
35 Maciejewski, H. (2013) Gene set analysis methods: statistical models and methodological differences. *Brief. Bioinform.* doi: 10.1093/bib/bbt002.
36 Nam, D. and Kim, S.-Y. (2008) Gene-set approach for expression pattern analysis. *Brief. Bioinform.*, **9**, 189–197.
37 Efron, B. and Tibshirani, R. (2007) On testing the significance of sets of genes. *Ann. Appl. Stat.*, **1**, 107–129.
38 Barry, W.T., Nobel, A.B., and Wright, F.A. (2005) Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics*, **21**, 1943–1949.
39 Newton, M.A., Quintana, F.A., den Boon, J.A., Sengupta, S., and Ahlquist, P. (2007) Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.*, **1**, 85–106.

4
Role of Modeling and Simulation in Toxicology Prediction

Antje-Christine Walz, Hans Peter Grimm, Christophe Meille, Antonello Caruso, Neil Parrott, and Thierry Lavé

4.1
Introduction

The consideration of drug metabolism and pharmacokinetics (DMPK) and physicochemical properties as part of the drug discovery and development process has substantially improved the quality of drug candidates with respect to pharmacokinetic properties. The main reason for attrition in development is now related to safety and efficacy, and the current challenge is to establish a relationship between pharmacokinetic and pharmacological/safety endpoints starting from early discovery stages. Mechanism-based pharmacokinetic/pharmacodynamic (PK/PD) modeling for both efficacy and safety assessment is being successfully employed throughout all discovery and drug development stages for data interpretation, study design, and enhanced decision making. Mechanism-based PK/PD models range from simple PK/PD models to physiologically based pharmacokinetic/pharmacodynamic (PBPK/PD) models, as well as more complex systems biology. Because the models are developed to describe the underlying biology and pathways, they provide the opportunity to separate system-specific parameters from compound-specific parameters [1,2]. They also enable extrapolation from *in vitro* to *in vivo* settings, from animals to humans, from healthy volunteers to a disease population, and from one compound to other similar compound. In addition, unlike empirical models, the model parameters in a mechanism-based PK/PD model can be experimentally measured and in some cases estimated *in silico*.

The specific set of questions to be addressed using PK/PD modeling depends on the stage of drug development. For late-stage preclinical and clinical development, PK/PD modeling is used to impact selection of dose regimens and more generally to support optimal study design for preclinical (e.g., toxicology) and clinical studies. For the discovery and early development stages, the understanding of the PK/PD relationship helps to better define the target profile and support the selection of lead compounds to advance into clinical studies based on the separation between efficacy and safety, the estimation of

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

safe and efficacious dosing schedules for animals and humans, and the prediction of safe starting doses for first-in-human studies. It also provides a basis for optimal design of proof-of-concept (PoC) studies (usually phase IIa) in relevant patient populations.

This chapter is focused on the application of predictive modeling and simulation approaches and more specifically of mechanism-based PK/PD modeling during discovery and early drug development supported by specific case studies as well as examples from the literature.

4.2 The Need to Bring PK and PD in Predictive Models Together

A vast amount of pharmacokinetic, pharmacological, and safety data is used to drive decision making for compounds to be progressed into development. Successful integration of discovery data is key to selecting promising compounds for development.

4.2.1 Physiologically Based Pharmacokinetic Modeling

We suggest integrating all available preclinical data into a single interpretable context by using physiologically based pharmacokinetic modeling, further integrating pharmacology and safety data into the model. The PBPK/PD approach can incorporate a diverse set of physicochemical, pharmacokinetic, and pharmacodynamic data and it is scientifically driven, relying on validated mechanistic models and thereby eliminating a large portion of the usual bias in data interpretation.

Such biological mechanism-based models built on the basis of physiologically relevant parameters provide the opportunity to translate *in silico*, *in vitro*, and *in vivo* preclinical data into knowledge that is relevant to the situation in humans.

As a simple example, take a measured value for the solubility of a compound. By itself, this data item tells us little about the absorption of the compound in humans. Only when combined with permeability, pK_a, and other physicochemical properties can one start to make predictions; that is, the first necessary step is data integration. And yet, without any model (be it physiology based or statistics based), the data make little sense; that is, the second step must be the construction of a suitable model. With appropriate data and a trustworthy model, predictions can be attempted. However, in order to gain additional confidence, comparison of model predictions and *in vivo* outcomes in animal species should be attempted. After this third step, predictions can be made with reasonable confidence.

As another example, consider an *in vivo* test fundamental for the evaluation of CNS compounds, namely, the induction of locomotor activity (LMA) or variants

thereof. *Per se*, a dose–response analysis of LMA in mice is of little value when it comes to predictions of efficacy against a psychiatric condition in humans. On the one side, LMA needs to be linked to receptor occupancy *in vitro* and to measured target affinity; this allows to account for species differences, for example, in protein binding and affinity. On the other side, the LMA needs to be related to a functional experiment in animals that in turn has been linked to the psychiatric disorder under investigation, possibly based on other compounds (e.g., competitor information in the public domain). Ultimately, this knowledge has to be combined with pharmacokinetics. Modeling is indispensable as the "glue" between these pieces. Furthermore, mechanistic/physiologically based models have the capability to capture strong inherent nonlinear processes that are often problematic in purely statistics-driven approaches.

4.2.2 Mathematical (PBPK, PK/PD) Modeling

Mathematical (PBPK, PK/PD) modeling enforces the explicit formulation of hypotheses. Based on a given set of hypotheses, predictions can be made and later – ideally – tested. Several hypotheses can coexist and the resulting predictions delimit the range of likely outcomes. In this way, models offer the possibility of explicitly incorporating uncertainty into the decision-making process and translating that uncertainty into a measure of confidence in the simulation and/or prediction. Unlike other empirical models, the assumption and the uncertainty in a mechanistic PK/PD model may be reduced by further experimental data collection. Quantifying key uncertainties and providing a range of possible outcomes based on the current knowledge of the PK properties of the compound will allow for more informed decision making. In addition, biologically based mechanistic models offer a platform for incorporating the known variability between different patient populations [3].

One of the most powerful applications of mechanism-based modeling and simulation is the integration of information from different studies, especially the findings from clinical studies, to advance the best compounds at preclinical development. Mechanism-based PK/PD models allow separation of system-specific parameters from compound-specific parameters; thus, the same system parameters from studies of other compounds in humans on the same pharmacology or disease process can be used to model similar compounds in the same system. This is particularly useful, considering the challenge of translating research from the animal model to human efficacy and safety.

4.2.3 Predictive Tools

In preclinical discovery and development, the key activity is to use "predictive tools" and "evaluation tools" to help screen out the candidates most likely to

have serious undesired side effects and identify those most likely to become safe and effective treatments. Because of the predictive nature of the modeling needed in preclinical development, such models ideally would be mechanistic as opposed to empirical. That is, the mechanistic model can provide better capacity for credible extrapolations. For example, *in silico* models based on large clinical safety database such as adverse drug reactions (ADRs) are now used in preclinical development [4]. The learning–confirming cycles of drug development as defined by Sheiner and Steimer [5] also apply to feeding the learnings from clinical data and clinical experience back into drug discovery to better design predictive models at the preclinical stage. There are vast amounts of clinical data from compounds with similar structure/class that can help in the selection of lead compounds or backup molecules at early development. Model-directed drug development processes should be bidirectional. Successful prediction and application of a preclinical model for phase 1 and early PoC clinical studies should not be seen as an end in itself. As compounds move through subsequent development stages, new learning from the clinical study of compounds should be incorporated into the model, and an updated model should be a better "predictive" model for the backup program or similar class of targets. In summary, the mechanistic PK/PD model provides a quantitative framework for translation and systematic integration of data and learning from clinical studies to preclinical development.

4.3 Methodological Aspects and Concepts

4.3.1 "Cascading" Drug Effects

The principle of cascading models to relate processes on the causal path between drug administration and response has been described recently [1,2]. With this approach, the effect of drugs from one process in the chain of events to the next can be described. With such an approach, pharmacokinetics can be used to estimate target exposure that can then be related to pharmacodynamics. Target effect considerations represent the initial steps of PK/PD model building, which can then be extended to include downstream effects such as molecular target activation, physiological and pathophysiological measures, and clinical ratings [1,2,6] as drug candidates move through the discovery and development process. The relationships between the various processes are solely dependent on the functioning of the biological system and are therefore independent of the drug.

Mechanism-based PK/PD modeling constructed on the basis of cascading models constitutes, therefore, the scientific basis for prediction of the ultimate clinical effects of novel drugs based on the response obtained at the various stages of the chain of biological events.

4.3.2 Linking Exposure and Effect

In many cases, PK predictions can be made with reasonable confidence for small molecules using, for example, physiologically based models [7]. Physiologically based models are useful to simulate exposure in plasma and in individual tissues, including target tissues. The target exposure can be linked to efficacy parameters measured *in vitro* in order to estimate receptor occupancy that provides the link between exposure and effect. Lack of solid understanding of target exposure can invalidate prior PK/PD modeling. Target exposure can be challenging to estimate when active transport processes are involved, for example, in tissue uptake. Recently, improved estimation of the transporter Michaelis–Menten parameters in the *in vitro* assay for quantitative predictions of transporter dynamics *in vivo* has opened the door for the possibility of incorporating transporter kinetics in a PBPK model for improved prediction of target tissue concentration [8].

Many biological responses can show a delay relative to drug concentrations in plasma, which is often referred to as hysteresis. When distribution to the site of action becomes rate limiting and determination of drug concentration at the target site is difficult, the effect compartment model (or so-called link model), for example, provides a useful way for estimating the concentration at the effect site [9]. The model assumes that drug in the effect compartment does not contribute to the pharmacokinetics of drug in plasma. The effect compartment modeling is useful when response delays are due to drug distribution from plasma to the effect site. The use of the effect compartment allows collapse of the hysteresis loop and estimation of effect concentration at the target site and subsequently leads to an improved estimation of *in vivo* potency of compounds.

4.3.3 Receptor Occupancy/Enzyme Inhibition

The potency (i.e., the EC_{50}) and the intrinsic activity (i.e., maximal effect, E_{max}) of a drug are functions of compound-specific (i.e., receptor affinity and intrinsic efficacy) and system-specific properties (i.e., the receptor density and the function relating receptor occupancy to pharmacological effect). Classical receptor theory explicitly separates drug-specific properties and system-specific properties as determinants of the drug concentration–effect relationship and therefore constitutes a theoretical basis for the prediction of this relationship. Not surprisingly, receptor theory is increasingly applied in mechanism-based PK/PD modeling to explain and predict (variability in) *in vivo* drug concentration–effect relationships [1,2].

4.3.4 Transduction into *In Vivo* Response

Transduction refers to the processes of the translation of the receptor activation into the ultimate pharmacological response. Specifically, the binding of a drug to

a biological target initiates a cascade of biochemical and/or electrophysiological events resulting in the observable biological response [1,2]. When transduction is fast (i.e., operating with rate constants in the range of milliseconds to seconds), relative to the rate constants governing the disposition processes (typically minutes to hours), the transduction process does not influence the time course of the drug effect relative to the plasma concentration. *In vivo* transduction can also be slow, operating with rate constants on the order of hours to days, in which case transduction becomes an important determinant of the time course of drug action [1,2].

4.3.4.1 Indirect Response Models

A widely used mechanism-based modeling approach for describing delayed biological response is the indirect response (IDR) model. The IDR model is based on the concept of homeostasis; that is, physiological entities (e.g., proteins or body temperature) are kept in dynamic equilibrium by balancing their buildup and their loss. Pharmacological activity can either stimulate or inhibit either of these processes as nicely reviewed by Mager *et al.* [10].

4.3.4.2 Transit Compartment Models

Models have been proposed in which transduction is modeled mechanistically on the basis of intermediary processes between pharmacokinetics and response. The so-called transit compartment model (TCM) has been proposed for this purpose. This model relies on a series of differential equations to describe the cascade of events between receptor activation and final response [11–13]. As with the IDR, traditional TCMs are motivated physiologically but are often phenomenological descriptions of pharmacodynamic response.

An interesting application of transduction models in drug development is the model for describing the time course of myelosuppression [14]. The model consists of a proliferating, drug-sensitive compartment, a series of transit compartments representing maturation, and a compartment of circulating blood cells. A key feature of this model is the separation of fixed system-specific parameters (such as proliferation and transit time) and a small number of drug-related parameters that have to be estimated. The model, "calibrated" with known compounds, can be used to translate the myelosuppression of new compounds from animals to humans. Clearly, the model is semi-phenomenological in the sense that it does not aim at profound understanding of the complex processes in blood cell maturation. It is rather tuned with a minimum of parameters required to capture the most important aspects of interindividual differences as well as drug action.

Transit compartment models are also used to characterize the antitumor effect in *in vitro* [13] and *in vivo* experiments [15,16]. The response of anticancer drugs is delayed relative to the time course of drug exposure. Semimechanistic models relate the tumor growth inhibition time course to the PK of the drug by separating system-specific (e.g., tumor size and growth rate) from drug-specific (e.g., potency and drug-induced kill rate) properties [15,16].

These models allow for the estimation of a concentration resulting in tumor stasis, which can be regarded as a reference concentration to be achieved for attaining significant anticancer activity in humans. Rocchetti *et al.* have shown a good correlation with the predicted concentration required for tumor stasis and the reported efficacious exposure for select chemotherapeutic agents in cancer patients [17]. Modeling the time course of anticancer effect as a function of the dosing protocol enables to predict intermittent dosing regimen reaching equivalent antitumor efficacy compared with chronic dosing.

4.3.5 Disease Modeling

One of the key challenges in early drug discovery and development, especially for novel first-in-class compounds, is that we often do not know whether a highly potent compound that inhibits specific receptor function or alters specific pathway bears any relevance to efficacy in the disease population (e.g., modulation of disease progression or symptomatic relief). Without any doubt, the effort in development of new compounds is directed toward drugs that can halt or alter disease progression. Therefore, it is important that PK/PD modeling can also be extended to include the effects on disease progression. Understanding disease progression is critical as the optimal pharmacodynamic response needs to be defined in the context of their effects on the processes of disease progression. In early drug discovery and development, compounds that target a specific pathway are often assumed to have impact on the ultimate disease process. For a complex disease, this target could be difficult to validate until a proof-of-concept study in a disease population is performed.

Disease progression modeling was listed in the Critical Path Initiative by the US FDA [18]. Development of new biomarkers for disease processes was identified as the highest priority for scientific effort and quantitative modeling of the disease process, incorporating what is known about biomarkers, would be an obvious next step. Disease progression modeling has been applied in modeling of clinical studies, such as Alzheimer's disease, Parkinson's disease, and viral dynamics in HIV- or HCV-infected patients [19,20].

In the context of antiviral drugs, PK/PD models have been developed to describe the effect of compounds on viral replication using empirical E_{max} models [21], as well as PK/PD disease models to link the viral and cell dynamics in patients as a function of time. Useful disease models for viral dynamic changes in patients have been reported recently for compounds targeting HIV and HCV [20,22,23]. The development of a HCV viral dynamic model to describe viral load changes in patients has helped understand the mechanism of the antiviral efficacy of interferon-alpha and ribavirin [21–23]. Subsequently, these initial models were extended to account for the presence of both wild-type virus and low level of telaprevir-resistant variants in estimating the required treatment duration to eradicate the virus [24].

The following section describes and illustrates with selected examples the application of predictive modeling and simulation approaches focusing on PBPK and PK/PD modeling for safety and efficacy along the discovery and development process.

4.4 Application During Lead Optimization

During the lead optimization phase, high-throughput chemistry generates numerous compounds and the physicochemical, pharmacokinetic, and pharmacological properties targeted within a particular project need to be defined. Since PK/PD models integrate all properties in a single framework, they can be extremely useful to define the range of properties needed to achieve a desired clinical outcome in terms of extent and duration of effect.

4.4.1 Example 1: PK/PD Modeling for Identifying the Therapeutic Window between an Efficacy and a Safety Response

This example demonstrates the use of PK/PD modeling in a lead optimization stage project to determine the relationship between plasma concentrations and both an efficacy and safety PD effect to determine which candidate compound has a large enough margin of safety for clinical success. Investigators have shown that selective agonists of the alpha 1a receptor can cause increased pressure in the urinary tract [25], which could lead to decreased symptoms from stress urinary incontinence (SUI). However, agonism of the alpha 1a receptor can also cause increased blood pressure (e.g., the alpha 1a agonist, midodrine, treats orthostatic hypotension, but can cause dangerously increased supine blood pressure if taken too soon before bedtime). Blood pressure is a carefully regulated physiological phenomenon, and, at low doses of an alpha 1a agonist, as the blood pressure increases the heart rate rapidly decreases to compensate, resulting in no net change in blood pressure. But with increasingly high doses of an alpha 1a agonist, the ability of the heart rate to compensate is overcome and the blood pressure will start to increase. The project described here was aimed at developing an alpha 1a partial agonist that could increase pressure in the urinary tract while avoiding any increase in blood pressure effects, which could be a promising medicine for stress urinary incontinence [26].

The first compound evaluated, referred to as compound 1, had been used in a clinical trial to demonstrate that this class of compounds has the potential to improve the symptoms of SUI [26]. Data pertinent to cardiac safety for compound 1 in humans (i.e., heart rate and supine systolic blood pressure) were available [23,26]. A preclinical model, the conscious minipig, showed increased urethral tone at low doses with candidate partial agonists as determined using a catheter that determined pressure throughout the length of the urethral tract.

With increasing doses of compound 1, the conscious minipig showed first a decreased heart rate and then eventually an increased blood pressure, as observed in humans [23,26]. To assess the predictability of the preclinical animal model, PK/PD modeling was used to compare cardiac safety of the compound in clinical development, compound 2, in both the minipig model and humans. Once the predictability of the model was verified, the model was used to assess the margin of safety in lead optimization candidates to determine which compounds to move forward into clinical development.

The PK/PD analysis began with plotting PD response as a function of plasma concentrations for individual minipigs. No hysteresis was observed for intraurethral pressure (IUP), blood pressure (BP), or heart rate (HR) indicating that the time course of the effect could be related to the time course of plasma concentrations. The data for multiple animals and dose levels were pooled to establish the PK/PD relationship. The following Hill equations were used to fit the observed data for PD effects (i.e., ΔIUP, ΔHR, and ΔBP, where the Δ symbol indicates change from baseline) as functions of plasma drug concentration (C_p):

$$\Delta\mathrm{IUP} = E_{\max,\mathrm{IUP}} \times C_p^{\gamma} / \left(\mathrm{EC}_{50,\mathrm{IUP}}^{\gamma} + C_p^{\gamma}\right),$$

$$\Delta\mathrm{BP} = E_{\max,\mathrm{BP}} \times C_p^{\gamma} / \left(\mathrm{EC}_{50,\mathrm{BP}}^{\gamma} + C_p^{\gamma}\right),$$

$$\Delta\mathrm{HR} = E_{\max,\mathrm{HR}} - \left(E_{\max,\mathrm{HR}} - E_{0,\mathrm{HR}}\right) \times C_p^{\gamma} / \left(\mathrm{EC}_{50,\mathrm{HR}}^{\gamma} + C_p^{\gamma}\right).$$

The resulting models were used to determine the plasma concentrations that would result in an efficacious response (i.e., a high enough value of ΔIUP to result in efficacy) or that would result in a 5 bpm drop in HR or a 5 mmHg increase in BP, which would be considered an unacceptable change in heart rate or blood pressure.

This PK/PD assessment was useful for the project for several reasons. First, the modeling allowed a clear assessment of compound efficacy and the safety window between the plasma concentrations required for efficacy and those that would result in cardiac safety issues. For compound 1, the safety margin was about ninefold, that is, the C_p resulting in a 5 mmHg ΔBP was about nine times higher than the C_p required for efficacy. For candidate compound 2, the safety margin was lower (5.9) than that for compound 1, but for all other candidates the margin of safety was higher. Second, the PK/PD modeling could be combined with a human PK extrapolation to provide a clear prediction on whether efficacy could be achieved in the clinic while avoiding the cardiac safety effect. For example, compound 3 had rapid clearance and initially there were doubts as to whether it would be a suitable clinical candidate. But using the PK/PD models coupled with a prediction of human PK for the compound, it was determined that the margin of safety was high enough that even with a large peak-to-trough ratio, compound 3 not only would be efficacious in the clinic, but would also have a large margin of safety. These parameters have been used to select compounds for preclinical development. Finally, the PK/PD modeling was used to

provide guidance for the design of some clinical studies. For this guidance, a PK model was combined with the PK/PD models to simulate certain dose regimens and dose levels to determine an optimal clinical trial design.

4.5 Application During Clinical Candidate Selection

PK/PD has a great potential to assist clinical candidate selection where numerous factors must be considered and data related to the PK and PD of a compound need to be combined and compared in a rational way. This potential is illustrated below with a number of examples. Examples are also available in the literature. For example, Parrott *et al.* [27] demonstrated the use of combined PBPK and PK/PD modeling to select the best clinical lead from among five candidates. The preclinical data for the five candidates were integrated and the efficacious human doses and associated exposures were estimated (Figure 4.1). The PBPK models were linked to a PD model so that the dose resulting in a 90% effect could be identified. This example showed that the PBPK approach facilitates a sound decision on the selection of the optimal molecule to be progressed by integrating the available information and focusing the attention onto the expected properties in humans. Importantly, the method can include estimates of variability and uncertainty in the predictions to allow decisions to be based on significant differences between the compounds.

In another example, a mechanism-based PK/PD model incorporating target-mediated binding and clearance of the antibody was developed for the candidate selection stage of backup molecules for omalizumab [28]. The challenge addressed by the modeling was to understand the relationship between dose, *in vitro* affinity, and *in vivo* efficacy in humans for the follow-up compounds. The PK/PD model was developed based on clinical data of omalizumab and *in vivo* efficacy was based on a surrogate marker of maximum reduction in free

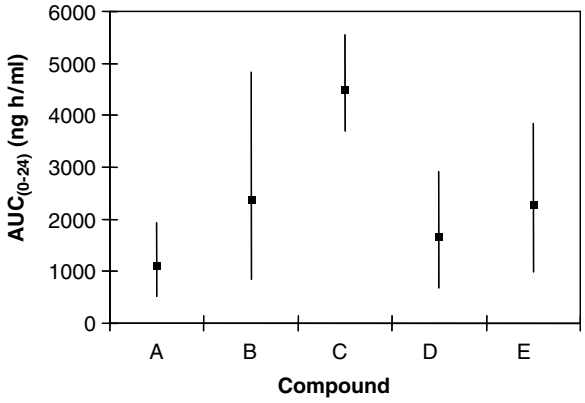


Figure 4.1 Simulations of exposures at the efficacious human doses for the five candidates.

IgE level. Unlike other empirical models, the construction of such a mechanism-based model makes it easy to separate the system/disease parameter from the drug parameter. As such, the model was flexible enough to account for different properties from another antibody, including different interactions of the antibody in the system including on/off rates from the antigen and capacity-limited distribution and elimination of antibody. A critical learning based on the sensitivity analysis of the model was that similar maximum reduction of free IgE could be achieved by half the dose if the antibody affinity for IgE was 5–10-fold higher than that for omalizumab. Further increase in affinity would not increase *in vivo* efficacy. This insight from the modeling could avoid expensive affinity maturation steps. The resulting model was able to help clinical lead selection of backup mAb. In addition, the model could be easily adapted for other follow-up compounds and used to model different disease populations (e.g., different baseline IgE levels).

4.5.1
Example 2: Translational PK/PD Modeling to Support Go/No Go Decisions

This example summarizes a translational PK/PD approach to rank compounds based on the anticipated therapeutic window in humans in order to select the most promising candidate for entry-into-human (EIH) testing. The aim of the translational PK/PD model was to project the QRS widening in humans at a concentration that is 30-fold higher than the expected human therapeutic exposure.

Five drug candidates, which were developed for the same therapeutic indication, were investigated for cardiovascular adverse effects after oral administration by gavage in beagle dogs. Telemetry systems were implemented to monitor cardiovascular parameters with minimal animal disturbance. Baseline subtraction and normalization by vehicle/placebo were employed to determine the pure drug effects. QRS widening was apparent for all five compounds. In order to quantitatively compare the compounds, a PK/PD model was built to relate the concentration–time course profile to the QRS data. In the second step, the concentration was corrected for the unbound fraction and used to simulate the QRS widening as a function of the free plasma concentration. This allowed to assess the anticipated therapeutic window in humans, which in turn supported the decision on testing in humans.

QRS data were modeled assuming a direct and saturating concentration–effect relationship. An E_{max} model with Hill coefficient was used:

$$\Delta\mathrm{QRS} = E_{\max} \cdot \frac{C_p^{\gamma}}{C_p^{\gamma} + C_{50}^{\gamma}},$$

where ΔQRS is the increase in QRS duration expressed as individual time-matched percent change from vehicle, E_{max} is the maximum effect, C_{50} is the concentration associated with half-maximal response, and γ is a slope parameter (Hill coefficient). Compartmental modeling was applied to describe the PK.

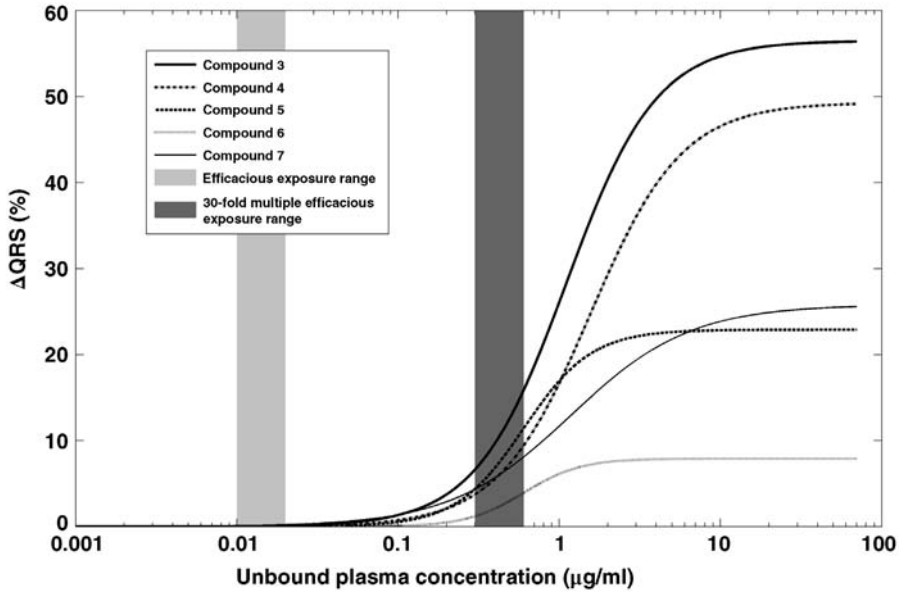


Figure 4.2 Simulated QRS widening effects of compounds 3–7 in the beagle dogs expressed as percent change from vehicle. Lines: unbound plasma concentration–QRS response relationships for the five candidates. Shaded areas: expected human efficacious concentration range and its 30-fold multiple.

As a result, the drug potency (C_{50}) in the dog was found to be similar when accounting for the differences in plasma protein binding. Conversely, the maximum response (E_{max}) was found to vary considerably between the drugs, suggesting differences in their intrinsic capacity to produce QRS widening [29]. By plotting the concentration–response curves for each drug candidate (after correction for interspecies differences in free fraction), it was possible to visualize the separation between QRS effects and human target exposures (Figure 4.2). The efficacious human exposure was based on first human PK prediction to cover EC_{90} (~100 nM or ~50 ng/ml) over 24 h and corrected for the unbound fraction. Significant QRS changes were apparent for all the tested compounds at a 30-fold multiple of the expected efficacious concentrations (corrected for the unbound fraction).

Assuming that the unbound plasma concentration–QRS response relationship is similar in the beagle dogs and humans [29], this PK/PD modeling approach enabled the prediction of QRS widening at selected concentrations that were of interest, namely, 30-fold the expected human therapeutic exposures. It showed that the therapeutic index of all tested compounds was less than the 30-fold multiple, which was considered as reasonable safety margin for humans. In conclusion, the translational PK/PD modeling approach led to the decision that none of the drug candidates were amenable to further development.

4.6
Entry-into-Human Preparation and Translational PK/PD Modeling

Translational PK/PD is a continuous process starting with human projections during preclinical development and stretching as far as into phase II. While compound selection is the focus at discovery stages, the attention for EIH shifts more to questions of safety and study design. In particular, regulatory toxicology studies (GLP-compliant; phase 0), single ascending dose/multiple ascending dose clinical trials (SAD/MAD, phase I), and PoC (phase II) studies are concerned.

4.6.1
Selection of Safe and Pharmacologically Active Dose for Anticancer Drugs

Phase I entry-in-human studies with anticancer drugs differ from other phase I studies in that they are mainly evaluated in patients whose disease condition is progressive and fatal. The FDA guideline for anticancer pharmaceuticals defines "the goal of selecting a starting dose is to identify a dose that is expected to have pharmacological effects and is reasonably safe" [30]. Translational PK/PD modeling is useful to improve the study design for EIH trials as illustrated in the examples below. Example 3 shows how biomarker information is collected and integrated into a model-based approach to project the pharmacological response in humans. Since most anticancer drugs have a narrow therapeutic window, the profiling of the therapeutic window in preclinical studies and the prediction to humans is a critical step. Example 4 illustrates how a translational PK/PD model for both safety and efficacy can be applied to select the most favorable dosing regimen in humans.

4.6.1.1 **Example 3**

Compound X is a monoclonal antibody developed to treat cancer patients. A PK/PD study was conducted and analyzed to predict the exposure–response relationship in the cross-reactive species. The proposed PKPD model was scaled to humans in order to support dose selection and to improve the study design.

Compound X is a noncompetitive antagonist that binds to the receptor Y. In the absence of drug, the endogenous ligand L binds to its receptor Y and triggers a signaling cascade that ultimately promotes tumor growth. When drug X is bound to the receptor, the ligand–receptor interaction is inhibited. The pharmacokinetics and pharmacodynamics of compound X were assessed in cynomolgus monkeys at various dose levels. Compound X showed a nonlinear PK profile suggesting a target-mediated elimination pathway. An increase in ligand L was observed after single and repeat administration of compound X and served as a PD marker to quantify the exposure–response relationship. Receptor-mediated elimination was assumed to be the major elimination pathway of the ligand, which is blocked when compound X binds to the receptor. The time course of the ligand kinetics in cynomolgus monkeys was modeled using a turnover model. In the absence of drug, a constant baseline level of the ligand is assumed

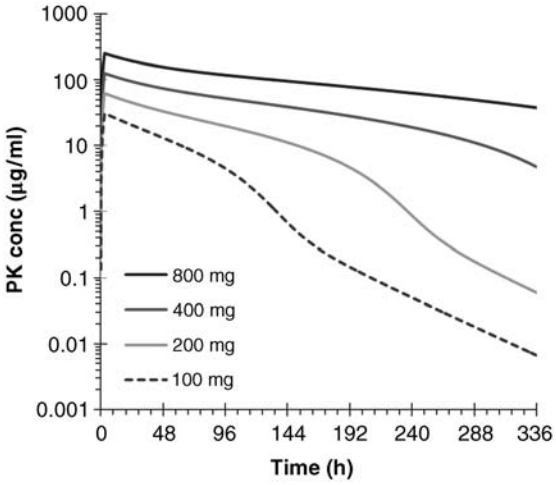


Figure 4.3 Simulated PK response in humans at different dose levels based on a translational PK/PD model built on monkey data.

due to equilibrium between the zero-order synthesis rate and the first-order loss rate (k_{out}). The drug action is assumed to block the receptor-mediated elimination of the ligand. In order to scale this pharmacological effect to humans, the drug-related parameters (IC_{50} and I_{max}) were assumed to be the same between monkeys and humans. For the ligand baseline level in humans, literature data were considered and the elimination rate of the ligand was scaled allometrically. For selection of a start dose, simulation of human PK (Figure 4.3) and PD (Figure 4.4) responses was performed.

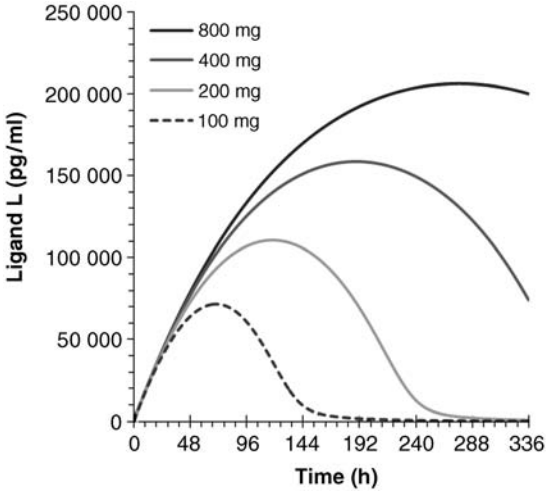


Figure 4.4 Simulated PD response in humans at different dose levels based on a translational PK/PD model built on monkey data.

The translational PK/PD model was used to design the EIH trial. Selection of the starting dose and the dosing regimen was supported by the model-based human PD response prediction. PK and PD sampling time points of the EIH trials were selected based on the simulated PK and PD profiles to ensure an informative study design, which allowed to quantitatively assess the PK/PD relationship in humans.

4.6.1.2 Example 4

Compound A is a targeted anticancer drug with a new mechanism of action. Antitumor clinical activity has been demonstrated with compound A; however, prolonged daily administration was poorly tolerated and resulted in treatment interruptions. Main adverse events in patients were neutropenia and thrombocytopenia. Compound B, a backup molecule that exhibits the same mechanism of action as compound A, was profiled in preclinical studies and thrombocytopenia was observed as dose-limiting toxicity. A translational PK/PD model was developed to describe the time course of the drug-induced thrombocytopenia by separating system-specific (e.g., platelet baseline level, maturation time, rate of progenitor production, feedback regulation) and drug-specific (e.g., drug potency) parameters. Human PK of compound B was predicted using a PBPK model and the time course of drug-induced thrombocytopenia in humans was simulated by accounting for system-specific differences while assuming drug potency to be the same between animals and humans [31]. For this simulation exercise, various dosing schedules were tested, which were predicted to provide similar anticancer effect. The efficacy projection was based on a semi-mechanistic PK/PD model that described the time course of drug-induced tumor growth inhibition in tumor-bearing xenograft mice. As a result, a dosing regimen of 5-day treatment followed by 23-day period of drug holiday per cycle was predicted to be most favorable in humans and was subsequently implemented in the EIH trial.

4.6.2 PK/PD for Toxicology Study Design and Evaluation

Among the different observations made in toxicology studies, often only few are amenable for PK/PD modeling and even less can be predicted beforehand to improve the toxicology study design. Therapeutic proteins and monoclonal antibodies in particular are an important exception to that rule, since toxicity is often driven by exaggerated pharmacology.

4.6.2.1 Example 5

For compound A, target-related toxicity was anticipated based on observations from different compounds sharing the same or similar mechanism of action. However, the signs of toxicity seen in the study appeared later and were much less pronounced than expected. Retrospective comparison with two similar compounds showed that (1) for compound B, which has ~10× higher affinity than

compound A, the time of appearance of toxicity seemed to decrease with dose (with a lag of ~2 months at the lowest dose); and (2) for compound C, which has more than 100× higher affinity than compound A, the time of appearance of toxicity did not depend on dose and was similar to that reported for the highest dose of compound B (1–2 weeks), suggesting that the appearance of this type of toxicity depends on the ratio of exposure and the dissociation constant at the target, that is, on target occupancy.

4.6.2.2 **Example 6**

Compound M, a monoclonal antibody, depletes its target from the blood. The original toxicology study was planned with a 13-week duration with weekly dosing at 100 mg/kg and a 13-week wash-out/recovery phase. With the available data from a single-dose PK study, in which the target was monitored and compared with existing models for similar compounds, it was possible to construct a turnover model for the target with a stimulated depletion by the compound. With this model, it was shown that the clinical observation had to extend until at least week 40 of the study to see partial recovery of the target, which was later confirmed. This showed furthermore that the slow recovery of the target is not due to any unexpected toxicity but is likely to be the normal result of pharmacological activity. This knowledge can be somewhat counterintuitive and would have been difficult to predict without quantitative modeling.

4.6.2.3 **Example 7**

Rich PK/PD data are usually available in telemetry studies monitoring cardiac safety. Accurate interpretation of such studies is complicated by interindividual variability of PK, interindividual variability of PD baseline values and PD susceptibility, circadian baseline variations, and time lag of the PD response with respect to PK. A PK/PD modeling approach was used in two recent projects with the benefit of objectively quantifying the exposure–response relationship. In this example, the compound was efficacious in an animal model of depression and pain. The compound was found to be a moderate inhibitor based on an *in vitro* hERG assay and estimated to have sufficient safety margin. However, the compound showed a tendency to prolong QT_c interval in a dog telemetry study. The challenge was to understand the mechanism of QT liability and to devise a strategy/screening assay for the backup compounds. There appeared to be some dissociation between concentration profile of parent compound (T_{max}: 1–2 h) and onset of QT effect (7–8 h postdose). A major metabolite was identified (30–40% of the parent in dogs), where the QT effect more closely coincided with the metabolite exposure profile.

The impacts of PK/PD modeling on the investigation into key mechanistic effects in this example can be seen from several aspects. Perhaps most significantly, the compartment modeling collapsed the hysteresis loop (effect versus concentration profile of parent compound) resulting in a profile more closely related to the concentration–time profile of the metabolite. Hence, this analysis

strongly supported the key hypothesis that the metabolite was responsible for the observed QT effect *in vivo*.

4.7 Justification of Starting Dose, Calculation of Safety Margins, and Support of Phase I Clinical Trial Design

Dose selection crucially depends on safety and efficacy projections. For example, the EMEA guideline regarding the first dose in humans [32] states "The estimation of the first dose in human is an important element to safeguard the safety of subjects participating in first-in-human studies. All available information has to be taken in consideration for the dose selection and this has to be made on a case-by-case basis." At the simplest level, this is done by projecting PK (using PBPK or other approaches) and relating C_{max} and AUC for safety (typically from the no observable adverse effect level (NOAEL), adjusted for species differences and with additional safety factors applied) or C_{min} and AUC for efficacy (typically inferred from an animal model). Even though quantities such as C_{min} for efficacy might be derived from PK/PD in animal models, this approach does not deserve the label of PK/PD since it falls short of many important aspects of drug action. PK/PD modeling using safety endpoints has been used successfully to analyze ECG alterations during telemetry studies in animals as described in the previous section.

With respect to compounds for which "factors of risk" have been identified according to the EMEA Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products [32], dose calculation should additionally rely on the MABEL (minimal anticipated biological effect level). Even without such known risks, health authorities have requested the calculation of MABEL in some recent examples of therapeutic proteins. The guideline specifies "The calculation of MABEL should utilize all *in vitro* and *in vivo* information available from pharmacokinetic/pharmacodynamic (PK/PD) data such as:

1) target binding and receptor occupancy studies *in vitro* in target cells from human and the relevant animal species;
2) concentration–response curves *in vitro* in target cells from human and the relevant animal species and dose/exposure–response *in vivo* in the relevant animal species.
3) exposures at pharmacological doses in the relevant animal species.

Wherever possible, the above data should be integrated in a PK/PD modelling approach for the determination of the MABEL."

Limited practical experience shows that a unique estimate of MABEL is often difficult to achieve due to disparities between, for example, *in vitro* receptor binding, *in vitro* cellular assays, and *in vivo* animal models. Taken to the letter,

that is, when the above-mentioned "factors of risk" have been identified, the lowest of these values is relevant and additional safety factors may be applied. In other cases, these estimates may seem exceedingly low considering the following limitations: technical (difficulty to provide formulations at very low doses, difficulty to assess exposure), ethical (if the study is done in patients, many patients are treated without likely benefit), operational (difficulty to recruit additional cohorts, especially if the study is done in patients), and economical (long duration of trials).

At this stage, all available preclinical (*in vitro* and *in vivo*) information is considered to estimate an efficacious dose for humans. If a PK/PD model can be established to relate exposure at the target site and target effects (as inferred from *in vitro* assays) in animal models, then this model can be used for translation to humans by adapting the species-specific factors such as metabolism, protein binding, receptor affinity, and so on. Ideally, if receptor occupancy as calculated from target organ exposure, protein binding, and receptor affinity adequately predicts the outcome of a behavioral test in animal models, there is a certain degree of confidence that a useful prediction of the efficacious dose in humans can be made by predicting receptor occupancy in humans. An obvious premise to do so is that the animal model in question can be considered relevant for the human disease.

The first hurdle to take is the estimation of the exposure at the target site. In the cases where the target compartment is in rapid exchange with circulating blood, it is assumed that the effect is driven by the unbound concentration of compound in the blood. This becomes critical in cases where the relevant pharmacological concentrations are impossible or difficult to assess either because of low penetration or because the pharmacological activity is driven by a (unknown) fraction of the compound present in the organ/tissue. This situation can be found when targeting the brain with small molecules (unknown free fraction) or with therapeutic proteins (unknown penetration), targeting tumors (limited by binding barrier, interstitial fluid pressure, "mechanical"), and targeting liver (unknown transport, metabolism, protein binding). Where this information is missing, mechanistic models have been built to estimate relevant exposure and possible delays of drug action and to identify limiting factors with large differences between species, phenotype, or influence of comedication.

At the target level, species differences of binding affinity at the target need to be taken into account. Furthermore, often only unbound compound confers pharmacological activity, and factors such as plasma protein binding have to be considered.

4.8 Outlook and Conclusions

In summary, the consideration of DMPK and physicochemical properties as part of the drug discovery and development process has substantially improved the quality of drug candidates, especially their pharmacokinetic properties. The

main reason for attrition in development is now related to safety and efficacy, and the current challenge is to establish a relationship between pharmacokinetic and pharmacological/safety endpoints starting from early discovery stages. Historically, empirical models were applied to address these questions. However, we suggest that more mechanism-based models are needed to support preclinical drug discovery and early development. In this chapter, several PK/PD models were presented that have been applied from drug discovery up to phase II. These mathematical models represent an important component of translational research, improving the efficiency of drug discovery and early development stages.

References

1 Danhof, M., de Jongh, J. *et al.* (2007) Mechanism-based pharmacokinetic–pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. *Annu. Rev. Pharmacol. Toxicol.*, **47**, 357–400.

2 Danhof, M., de Lange, E. *et al.* (2008) Mechanism-based pharmacokinetic–pharmacodynamic (PK–PD) modeling in translational drug research. *Trends Pharmacol. Sci.*, **29**, 186–191.

3 Rostami-Hodjegan, A. and Tucker, G. (2007) Simulation and prediction of *in vivo* drug metabolism in human populations from *in vitro* data. *Nat. Rev. Drug Discov.*, **6** (2), 140–148.

4 Krejsa, C., Horvath, D. *et al.* (2003) Predicting ADME properties and side effects: the BioPrint approach. *Curr. Opin. Drug Discov. Dev.*, **6** (4), 470–480.

5 Sheiner, L. and Steimer, J. (2000) Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu. Rev. Pharmacol. Toxicol.*, **40**, 67–95.

6 Mager, D., Wyska, E. *et al.* (2003) Diversity of mechanism-based pharmacodynamic models. *Drug Metab. Dispos.*, **31** (5), 510–518.

7 Jones, H., Parrott, N. *et al.* (2006) A novel strategy for physiologically based predictions of human pharmacokinetics. *Clin. Pharmacokinet.*, **45**, 511–542.

8 Poirier, A., Lave, T. *et al.* (2008) Design, data analysis and simulation of *in vitro* drug transport kinetic experiments using a mechanistic *in vitro* model. *Drug Metab. Dispos.*, **36**, 2434–2444.

9 Sheiner, L., Stanski, D. *et al.* (1979) Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to D-tubocurarine. *Clin. Pharmacol. Ther.*, **25** (3), 358–371.

10 Mager, D.E., Wyska, E., and Jusko, W.J. (2003) Diversity of mechanism-based pharmacodynamic models. *Drug Metab. Dispos.*, **31** (5), 510–518.

11 Sun, Y. and Jusko, W. (1998) Transit compartments versus gamma distribution function to model signal transduction processes in pharmacodynamics. *J. Pharm. Sci.*, **87** (6), 732–737.

12 Mager, D. and Jusko, W. (2001) Pharmacodynamic modeling of time-dependent transduction systems. *Clin. Pharmacol. Ther.*, **70** (3), 210–216.

13 Lobo, E. and Balthasar, J. (2002) Pharmacodynamic modeling of chemotherapeutic effects: application of a transit compartment model to characterize methotrexate effects *in vitro*. *AAPS Pharm. Sci.*, **4** (4), E42.

14 Friberg, L., Henningsson, A. *et al.* (2002) Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J. Clin. Oncol.*, **20** (24), 4713–4721.

15 Simeoni, M., Magni, P., Cammia, C., De Nicolao, G., Croci, V., Pesenti, E., Germani, M., Poggesi, I., and Rocchetti, M. (2004) Predictive pharmacokinetic–pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Res.*, **64**, 1094–1101.

16 Yang, J., Mager, D.E., and Straubinger, R.M. (2010) Comparison of two pharmacodynamic transduction models for the analysis of tumor therapeutic responses in model systems. *AAPS J.*, **12** (1), 1–10.
17 Rocchetti, M., Poggesi, I., Germani, M., Fiorentini, F., Pellizzoni, C., Zugnoni, P., Pesenti, E., Simeoni, M., and De Nicolao, G. (2005) A pharmacokinetic–pharmacodynamic model for predicting tumour growth inhibition in mice: a useful tool in oncology drug development. *Basic Clin. Pharmacol. Toxicol.*, **96**, 265–268.
18 Woodcock, J. and Woosley, R. (2008) The FDA critical path initiative and its influence on new drug development. *Annu. Rev. Med.*, **59**, 1–12.
19 Chan, P., Nutt, J. *et al.* (2007) Levodopa slows progression of Parkinson's disease: external validation by clinical trial simulation. *Pharm. Res.*, **24** (4), 791–802.
20 Bonhoeffer, S., May, R. *et al.* (1997) Virus dynamics and drug therapy. *Proc. Natl. Acad. Sci. USA*, **94** (13), 6971–6976.
21 Rosario, M., Jacqmin, P. *et al.* (2005) A pharmacokinetic–pharmacodynamic disease model to predict *in vivo* antiviral activity of maraviroc. *Clin. Pharmacol. Ther.*, **78** (5), 508–519.
22 Neumann, A., Lam, N. *et al.* (1998) Hepatitis C viral dynamics *in vivo* and the antiviral efficacy of interferon-alpha therapy. *Science*, **282** (5386), 103–107.
23 Dixit, N., Layden-Almer, J. *et al.* (2004) Modelling how ribavirin improves interferon response rates in hepatitis C virus infection. *Nature*, **432** (7019), 922–924.
24 Lawitz, E., Rodriguez-Torres, M., Muir, A.J., Kieffer, T.L., McNair, L., Khunvichai, A., and McHutchison, J.G. (2008) Antiviral effects and safety of telaprevir, peginterferon alfa-2a, and ribavirin for 28 days in hepatitis C patients. *J Hepatol.*, **49** (2), 163–169.
25 Blue, D., Daniels, D. *et al.* (2004) Pharmacological characteristics of Ro 115-1240, a selective alpha1A/1L-adrenoceptor partial agonist: a potential therapy for stress urinary incontinence. *BJU Int.*, **93** (1), 162–170.
26 Musselman, D., Ford, A. *et al.* (2004) A randomized crossover study to evaluate Ro 115-1240, a selective alpha1A/1L-adrenoceptor partial agonist in women with stress urinary incontinence. *BJU Int.*, **93** (1), 78–83.
27 Parrott, N., Jones, H. *et al.* (2005) Application of full physiological models for pharmaceutical drug candidate selection and extrapolation of pharmacokinetics to man. *Basic Clin. Pharmacol. Toxicol.*, **96**, 193–199.
28 Agoram, B., Martin, S. *et al.* (2007) The role of mechanism-based pharmacokinetic–pharmacodynamic (PK–PD) modelling in translational research of biologics. *Drug Discov. Today*, **12** (23–24), 1018–1024.
29 Mager, D.E. and Jusko, W.J. (2008) Development of translational pharmacokinetic–pharmacodynamic models. *Clin. Pharmacol. Ther.*, **83** (6), 909–912.
30 Food and Drug Administration (2010) Guidance for Industry: S9 Nonclinical Evaluation for Anticancer Pharmaceuticals. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). Available at http://www.fda.gov/downloads/Drugs/.../Guidances/ucm085389.pdf (last assessed March 2014).
31 Meille, C., Reddy, M.B., Walz, A., Retout, S., Nichols, G., Glenn, K., Pignatello, M., Bottino, D., Zhi, J., Middleton, S., and Lavé, T. (2012) Modified model of drug induced thrombocytopenia efficiently projects safe starting dose in human from preclinical data. Poster presentation, PAGE Meeting 2012. Available at http://www.page-meeting.org/?abstract=2446 (last accessed June 24, 2014).
32 EMA (2009) Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002988.pdf (last accessed June 24, 2014).

5
Genomic Applications for Assessing Toxicities of Liver and Kidney Injury

Philip Hewitt and Esther Johann

5.1
Introduction

5.1.1
Toxicogenomics in Drug Development

The scientific field of classical toxicology investigates the potential adverse effects resulting from exposure to chemical, biological, or physical agents. In drug development, these effects are evaluated by various short- and long-term exposures in different species and multiple endpoints, including organ weights or biochemical or histopathological alterations. These harbor inherent disadvantages, such as being time and cost intensive, as well as limited extrapolation power to humans, thus forming the chokepoint in drug discovery and development processes. In addition, these current observations provide only limited information about the underlying mechanism of a drug's toxicity.

The complete understanding of the adverse effects a drug may cause can only be achieved with the knowledge of its mode of action. It is expected that drugs can interact with a variety of molecules in an organism or specific tissue, including hitting various pathways that may cause alterations in gene expression levels and the subsequent modifications of the translated protein. These alterations finally lead to a pathological outcome, which can be either reversible or not.

These adverse response molecular mechanisms are much more sensitive compared with the classical endpoints of pathological observations. Furthermore, it has to be emphasized that changes in gene expression levels caused by a drug do appear much earlier, meaning that for toxicological investigations, fewer doses and treatment times may be needed. Together with the development of innovative technologies, as well as the increasing knowledge in genomics, these approaches raise the hope for a much improved, cost-effective and time-saving, and thus more efficient drug development process (Figure 5.1).

The combination of gene expression profiling technologies and toxicology has led to the scientific discipline of toxicogenomics [1,2]. Toxicogenomic approaches allow a more complete understanding of the underlying mechanisms

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

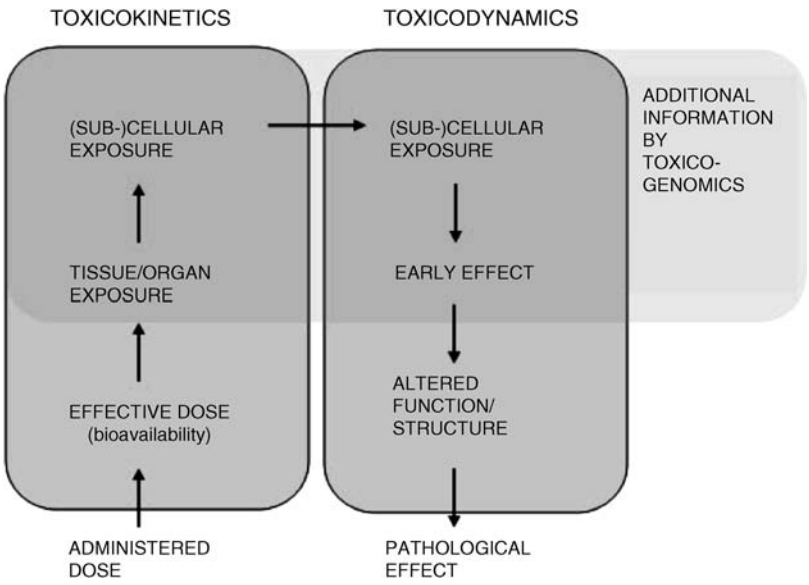


Figure 5.1 Network of toxicokinetics (uptake, distribution, and metabolism of a drug) and toxicodynamics (drug- or metabolite-induced responses and changes). Toxicogenomics may help to get deeper insights into the variabilities within this network on different levels.

of both pharmacology and toxicology than the classical toxicological assessment strategies.

These new technologies, when taken together with new bioinformatics tools that have been developed to analyze the flood of gene expression data, allow the comparison of thousands of genes. This results in the gathering of extraordinarily large amounts of data in just one single experiment, for example, the comparison of treated and untreated groups, thus providing an integrated retrieval of all cellular responses to drug exposure. Many researchers have been working on comparing expression profiles of drug candidates with those of toxicologically well-characterized reference compounds to classify new drug candidates and help prioritize the most promising structures [3].

One key drawback of measuring expression levels of genes alone is that no simple extrapolation to the levels of the encoded proteins is possible, since the "one gene–one protein" hypothesis is outdated. The mRNA of an expressed gene does not necessarily result in one protein due to a variety of modifications at different levels of processing that may lead to proteins with various functions or even a lack of transcription by silencing RNA (siRNA) or microRNA (miRNA). Therefore, the field of toxicogenomics is inseparably linked to other omics disciplines, focusing on what happens after expression, namely, proteomics and metabol(n)omics (Figure 5.2). Proteomics, for example, deals with the separation, quantitation, functional characterization, and regulation of all aspects of expressed proteins. Metabol(n)omics deals with physiological profiling and

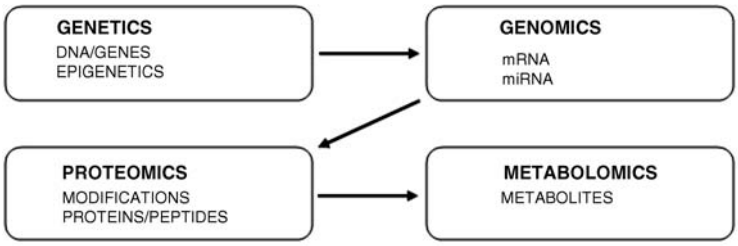


Figure 5.2 Overview of the different molecular levels of omics technologies and the strong linkage between these.

the metabolome can be considered to be the collection of all endogenous metabolites in a biological cell, tissue, organ, or organism, which are the end products of all cellular processes.

These technologies have great utility in preclinical test strategies since repeated administrations of low doses, mimicking the therapeutic range of a new drug, as well as of high doses for the distinct identification of toxic effects are tested in early phases of drug development. Thus, genomic (transcriptomic, proteomic, and metabolomic) data of low-dose applications provide a promising tool for the detection of the relationship between changes in gene expression and a pathological outcome. Any alterations of specific, key molecules either at the messenger RNA (mRNA) level or at the protein level can then be linked to the toxicological and/or pharmacological endpoints observed in classical *in vitro* and *in vivo* studies. By the comprehensive evaluation of these large amounts of data, hypotheses about the mode of action can be generated.

Besides the enormous advantages provided by these different genomic approaches, these are also challenges in utilizing these types of data, and will be addressed later on in this chapter.

In the following sections, the different mRNA expression array methods are introduced and subsequently the feasibilities and limitations of these technologies are discussed for the assessment of adverse effects in two of the major affected organs due to drug exposure, namely, the liver and kidney.

5.2
Toxicogenomic Approaches

5.2.1
High-Throughput Expression Profiles and DNA Microarrays

The increasing availability of whole-genome sequence information from various species has allowed the development of a multitude of DNA microarrays. These arrays enable one to quantitatively compare the transcriptional activity of potentially tens of thousands of genes between any biological samples, for example,

drug treated versus untreated. Nowadays, a variety of vendors offer microarray platforms, and although different technical platforms exist, they all follow similar basic principles of array technology. The basic setup of a DNA microarray is a small, solid chip, onto which numerous (hundreds to tens of thousands) oligonucleotides or copy DNAs (cDNAs) are covalently bound. Each nucleotide sequence thus represents an individual gene and is placed at a defined position on the chip. Isolated RNA, either mRNA or total RNA, from the tissue and species of interest can then be converted to cDNA by reverse transcription, fluorescently labeled, and hybridized to the microarray. Despite all of the potential problems associated with sample isolation and preparation, there are also other aspects, which may affect gene expression data. For instance, the experimental conditions under which the model systems (*in vivo* and *in vitro*) are kept are of importance. Factors such as age, gender, feeding state, or medium composition have an influence on gene expression levels. It is important to consider all these sources of variation when performing gene expression profiling.

Detailed descriptions of these techniques and their applications are given elsewhere [4–8].

5.2.2 Data Analysis

The successful application of toxicogenomics for the better understanding of organ toxicities highly depends on proper data analysis. Regarding the tens of thousands of data points generated by one high-density microarray, the major challenge is using the correct analysis. A variety of different bioinformatics tools exist and generally the analysis of such large amounts of data can be done by using several methods. It is possible to not only detect the genes related to substance-induced toxicity, but also perform class comparisons of different expression profiles induced by different classes of toxicants. For the prediction of possible hepatotoxic effects, these applications are important and the use of training databases constructed from gene expression profiles of known toxicants is needed. Training sets can be created by microarray analysis of tissues exposed to various toxicants with known toxic profiles compared with untreated controls. New and unknown compounds can then be tested in the same tissues and the expression patterns compared. Therefore, it is important to have rigid/controlled experimental conditions, since these may have a significant influence on the gene expression changes and subsequently any potential classification. The so-called unsupervised techniques do not need previously generated data of known hepatotoxicants. These techniques are used to find expression patterns and relationships in data sets without classifying them.

One of these techniques is called principal component analysis (PCA), an approach used to reduce the dimensionality of these complex data sets [9]. Generally, three components can be used to visualize samples on three-dimensional plots and similarities of gene expression between different biological samples observed as clusters. A very common approach is the hierarchical

clustering method, which can be very useful to characterize the number of classes of genes represented by the treatments (class discovery).

An example of a bioinformatics technique widely used for the prediction of complex data sets is the "support vector machine" (SVM) algorithm. SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns, and are used for classification and regression analysis. A SVM training algorithm can build models that then classify new samples into one category or another, already defined by the user.

Another approach is given by relevance networks, which analyze microarray data at the genome level by comparing genes in a pairwise manner and subsequently calculating correlation coefficients and threshold values. Pairs above the threshold values are then kept and posed as nodes. Thus, a gene could be directly or indirectly linked to other genes as well as to phenotypes. Combining these different techniques for the analysis of high-quality data will identify specific expression patterns of classes and also the classification of unknown compounds.

To aid in the proper evaluation of these new techniques, the Food and Drug Administration (FDA) set up the MAQC (MicroArray Quality Control) Consortium. The first phase (MAQC-I) was set up to provide QC tools to avoid procedural failures, as well as to provide guidelines for microarray data analysis. In addition, the advantages and disadvantages of the various platforms and analysis methods were evaluated [10]. The second phase of the MAQC project (MAQC-II) aimed to assess methods in developing and validating microarray-based predictive models. "Best practices" for such predictive models were established and published a few years later [11].

5.3 Specific Applications of Toxicogenomics

5.3.1 Mechanistic Toxicogenomics and Risk Assessment

The knowledge of how a compound exerts its pharmacological and toxicological effects and how to distinguish between these is of major importance in drug development. There exist many difficulties for revealing relevant mechanistic information, for example, the determination of time and doses at which changes are meaningful. In addition, it is important to be aware that for a given time point more than one relevant mode of action might be present (on-target, off-target) and it should also be kept in mind that one expression profile only resembles a snapshot in time within the cell/organ/organism. The elucidation of a compound's mechanism can also be used to predict the toxicity of a compound, being possible by the identification of signatures caused by model toxicants and comparing expression patterns of unknown compounds. Therefore, it is important to generate these signatures based on an extensive training set of

model compounds with well-defined outcomes. The mechanistic understanding of a new drug's toxic liability has become more important for the risk assessment process. By working together with pathology and toxicology, the changes in expression levels can be used to create a meaningful context to help explain any pathological changes observed.

5.3.2 Toxicogenomic Profiling of Hepatotoxicity

5.3.2.1 Hepatotoxicity in Drug Development

The liver is the major organ of endogenous metabolism and it is also responsible for the metabolism of xenobiotics and facilitates the detoxification of toxicants. Therefore, the liver itself is often more exposed to xenobiotics than other organs and remains one of the main targets of many potential adverse effects. Drug-induced hepatotoxicity is a common issue in drug development and frequently a cause of liver injury, and accounts for a large proportion of acute liver failures worldwide. In addition, drug-induced idiosyncratic hepatotoxicity is the most common cause of postmarketing drug withdrawal, emphasizing that it is critical to identify potential signals of hepatotoxicity early in preclinical and early clinical trials during drug development.

Due to this sensitivity to adverse drug reactions, toxicogenomic approaches provide a useful tool for the early and sensitive detection of changes in gene expression caused by drug candidates prior to pathological manifestations, at subtoxic doses. The investigation and understanding of how drug-induced hepatotoxicity develops on a molecular level can now be performed in a routine and well-controlled manner.

The fast emerging platforms of commercial vendors also allow the comparison of expression responses to xenobiotic exposure across species, thus providing a more comprehensive analysis, which is important for toxicological assessments, as animals (e.g., rodents and dogs) are used. When it comes to the liver, the hepatic transcriptome is complex and functional genomic approaches have led to the identification of issues in investigating the liver and its drug-induced pathogeneses with focus on the molecular pathways subsequently leading to hepatotoxicity. These investigations also help clinicians to better adapt therapies, since different phenotypes may have a greater or lower risk for adverse effects.

In the following sections, the functional and structural properties of the liver are briefly discussed and toxicogenomic approaches are summarized subsequently.

5.3.3 Functional and Structural Properties of the Liver

The overall conception is that changes in gene expression profiles may indicate toxicity at an early stage due to xenobiotic-mediated changes in gene expression levels, which are mostly detectable prior to pathological alterations [12]. Second, the detection of gene expression changes occurs not only at toxic doses but also

at subtoxic doses, which has the potential advantage of identifying those compounds that cause chronic hepatotoxicity. Toxicogenomic investigations concerning hepatotoxicity have been in focus, since the liver is the main organ of xenobiotic metabolism.

In addition to the complexity and variety of functions of the liver, its architecture is unique, macroscopically divided into lobes and lobules that appear to have the same histological morphology consisting of multiple units defined by a central vein, surrounded by four to six portal areas [13]. These structural units consist of up to 15 different cell types and differing functional properties depending on the localization within the lobule. Keeping these fundamental properties in mind is essential for the understanding of liver function as well as of response processes leading to toxic effects. Due to these unique properties, the liver transcriptome is not yet fully understood and shows high complexity regarding the fact that it includes up to 40% of the approximate 50 000 mammalian genes [14], thus making toxicogenomic approaches complicated. Many interactions may be context specific, depending on different components that are active at certain cellular stages, polymorphisms, and/or the present composition of cell types. In addition, the transcriptome of the liver can double or triple in a diseased state or a state of drug response.

Exogenous factors such as age, gender, nutritional state, or alcohol intake can have a remarkable impact on the hepatic gene expression. All these factors may have an influence on genomic data and therefore pose as possible confounding factors that should be taken into account when analyzing gene expression data.

5.3.4
Liver Morphology

The blood supply is provided by two main vessels, hepatic artery and portal vein, and the blood flow to or from the lobe can underlie individual variations in different species [15]. Differences in the primary source of blood flow have an influence on the hepatic response to drug exposure [16]. For instance, the variation of localization of the blood flow may have an impact on the distribution of nutrients, xenobiotics, and their metabolites within the liver lobes, which may lead to lobe variations [17–19]. An example for this is the lobe-specific accumulation of heavy metals during diseases and liver development [20–22] as well as the lobe-specific differences in the development of liver cirrhosis [23], and the differential response to several carcinogens has also been reported [24].

This has also been shown using toxicogenomics, where transcriptional profiling of controls and acetaminophen (APAP)-treated rats also elucidated a clear lobe–lobe variation [25]. Malarkey *et al.* [26,27] could also detect a lobe variation in the severity of APAP-induced necrosis. This clearly shows that the tissue areas from which the samples for gene expression profiling are taken could have a significant influence on the data interpretation and potentially generate false conclusions.

5.3.5
Cell Types

For the analysis of hepatogenomic data, the proper understanding of the different cell types present in the liver is essential when considering the fact that each cellular compartment has a different gene expression. Furthermore, the cell composition can alter during a diseased state, thus having an impact on the transcriptome. There are at least 15 different cell types present in the liver [26,27], which can be separated into two main groups: parenchymal and nonparenchymal cells. Hepatocytes display the parenchymal part and represent about 60% of the total liver cells, whereas a variety of nonparenchymal cells are present having different functionalities, for example, Kupffer cells, hepatic stellate cells, or sinusoidal endothelial cells. All these different cell types influence the transcriptome depending on the different proportions during healthy or diseased states, thus having an influence on genomic data.

5.3.6
Functional Gradients

Due to the unique architecture of the liver, it has to be taken into account that functional gradients exist, which may have an impact on genomic data [16,28]. Most xenobiotics are metabolized to either detoxified or active metabolites, most commonly by cytochrome P450s (CYPs), of which a large number of isoenzymes exist. Jungermann and Katz [29] could determine that CYP expression levels vary within the microstructure of the liver, with a higher preference of CYPs in centrilobular hepatocytes. This may explain the regionally differing susceptibilities to hepatotoxicity in response to certain drug treatments.

There are also differences in certain physiological properties, such as oxygen saturation and metabolic activity [30], which can have an impact on gene expression profiles. In addition, it has also been reported that number, size, and metabolic activity of nonparenchymal cells vary regionally and again can influence any transcriptional changes.

5.4
Toxicogenomic Applications for the Better Understanding of Hepatotoxicity

5.4.1
Mechanistic Toxicology

In response to injury, the changes at the mRNA and protein levels precede the response at the physiological level. The mechanisms leading to toxic endpoints at the physiological level may be reconstructed by the examination of gene expression changes [31–33]. The early toxicogenomic studies demonstrated its value for the elucidation of the mechanisms of response to toxicants [34–37].

For example, Burczynski *et al.* [38] distinguished two mechanistically unrelated classes, cytotoxic anti-inflammatory drugs and genotoxic agents, based on a cluster-type analysis of expression patterns in HepG2 cells exposed to various compounds, thus establishing an approach for toxicogenomics-based discrimination of mechanisms.

The question whether genotoxic carcinogens lead to deregulation of genes involved in common pathways at an early time point of drug exposure was investigated by Ellinger-Ziegelbauer *et al.* [39] in order to get a better insight into mechanisms leading to tumor formation in the liver. Here, rats were exposed to four different carcinogens and all four compounds led to responses to DNA damage, detoxification, and survival/proliferation pathways. In addition, necrotic inflammatory response at the gene expression level correlated with histological findings of necrosis and inflammation for two compounds. Thus, the identification of a particular combination of mechanisms for the characterization of the early response to this class of compounds could be detected.

McMillian *et al.* [40] investigated drug-induced oxidative stress in the rat liver by using toxicogenomic approaches. They established expression signatures of macrophage activation using a training compound set. Subsequently, these expression patterns were used to identify macrophage activator (MA)-like compounds, such as CCl_4, gadolinium, coumarin, and several nonsteroidal anti-inflammatory drugs (NSAIDs).

By using this strategy, expression patterns for peroxisome proliferators (PPs) were also investigated. For the discrimination of expression signatures of oxidative stress, a large number of training compounds were used, since oxidative stress is induced by different pathways and also many different forms exist (reactive metabolites, redox cycling, glutathione (GSH) depletion, phase II induction, etc.). In addition, these authors reported that many genes involved in the response to oxidative stress were regulated by one single transcription factor, nuclear factor-like 2 (Nrf2), which binds to the antioxidant response element (ARE). ARE is known to play an important role in the induction of phase enzymes as well as in the management of oxidative stress [41]. By the use of these oxidative stress signature genes, the macrophage activator-like compounds clustered together, but away from the oxidative stress signature. Correspondingly, oxidative stress samples were discriminated from most other samples by using a set of six Nrf2-inducible genes, whereas MA samples generated different expression patterns and peroxisome proliferators were close to the control samples. Using the MA set, some PP-treated samples clustered together, but away from MA-treated samples, mostly due to opposite regulation of genes involved in β-oxidation of fatty acids.

Taken together, toxicogenomic approaches provide important insights into the possible mechanisms involved in the toxic pathways of hepatotoxicants, although additional studies are needed to better define and understand these mechanisms of toxicity.

Several other publications have confirmed the value of toxicogenomics to confirm various mechanisms of liver toxicity – especially in the rat. These include

mechanisms related to changes in fatty acid β-oxidation and lipid metabolism, peroxisome proliferation, oxidative stress and glutathione depletion, disruption of mitochondrial homeostasis, or hypoxia [42–46]. Blomme *et al.* [47] worked on understanding the mechanism underlying drug-induced microvesicular steatosis caused by cyclopropane carboxylic acid (CPCA) using transcriptomic techniques. In the liver of rats treated with CPCA, but not the butyrate derivative, genes encoding enzymes involved in mitochondrial β-oxidation, including acyl-coenzyme A, dehydrogenase, and carnitine palmitoyltransferase I, were strongly downregulated. CPCA treatment upregulated several genes involved in apoptosis, for example, c-myc, cytochrome *c*, and Bax. These data confirmed previous studies that demonstrated that mitochondrial damage is an important mechanism for the development of microvesicular steatosis [48,49].

Suter *et al.* [50] generated microarray data of Ro 65-7199-treated rats. Ro 65-7199 is a serotonin receptor 6 (5HT-6 receptor (5-hydroxytryptamine 6 receptor)) agonist used for the treatment of cognitive disorders. This receptor is mainly expressed in the brain, but has also led to fatty liver changes in rats. Microarray data revealed a downregulation of genes involved in lipid homeostasis and an upregulation of CYP2B2. Using the same compound, investigations have been made in an *in vitro* approach using primary rat hepatocytes [42]. They also observed a downregulation of genes involved in β-oxidation and upregulated expression of CYP2B2. Thus, this is a good example for how toxicogenomics can be used for discovery research; in particular, it may provide a basis for screening approaches in preclinical drug development.

5.4.2 Class Identification

The use of toxicogenomics leads to the possibility for a more advanced understanding of how drugs modulate gene expression levels, thus leading to organ toxicities. Structurally differing compounds may induce the same response pathways leading to similar toxic endpoints. Based on this principle, intensive investigations are being made using expression profiles of chemically different compounds and comparing these signatures according to their toxicological endpoints. For example, Hamadeh *et al.* [34,36,37] were one of the first groups to try this approach and focused on a rat model. Different chemical classes were included and cluster analysis demonstrated a close similarity of expression patterns for peroxisome proliferator inducers, but which clustered away from the pattern of the cytochrome P450 inducer, phenobarbital. This suggests that it might be possible to implement these approaches for the classification of novel compounds with other specific hepatotoxicity mechanisms. These authors also addressed the question whether it is possible to distinguish unknown compounds by comparison of the two chemical classes and investigated blinded liver samples of Sprague Dawley rats treated with chemicals and four known compounds as a training set. Data analysis revealed that 22 genes clearly exhibited different expression patterns between the two classes of compounds. With this

approach, they could successfully identify the classes of 12 out of 13 blinded samples. McMillian *et al.* [51] used this approach and classified three classes of hepatotoxicants successfully by their gene expression signatures. These studies demonstrated the possibility of both the prediction of properties using gene profiling and the prediction of the class of unknown compounds by using toxicogenomic methods.

5.4.3 Predictive Toxicology

This approach assumes that similar treatments leading to the same toxic endpoints may show the same changes in gene expression. Therefore, compound groups or classes (classified by properties such as structure, toxic endpoint, mechanism of action, or target) may induce characteristic expression patterns leading to the assumption that toxicogenomic approaches are able to predict toxic effects of unknown compounds. A strong correlation between classical endpoints of toxicity and gene expression changes has been reported [52] and additionally cDNA microarray analysis confirmed the observed histopathological findings such as necrosis and hyperplasia in a dose-dependent manner [34,36,37].

Ruepp *et al.* [53] constructed a large database with liver gene expression profiles from vehicle and treated samples, and a SVM algorithm was used to generate classification "rules." Using this approach, toxicity was detectable at an early stage of drug response before any other toxicological changes occurred. Based on histopathology and clinical chemistry, five classes were identified:

- Controls.
- Direct acting.
- Cholestasis.
- Steatosis.
- Peroxisome proliferation.

Using the SVM-based model, Ruepp *et al.* [53] identified a potential hepatotoxic response of tacrine after 6 h from gene expression profiles prior to pathological findings. But toxicogenomic assessment did not reveal toxicity after 24 h, which implies the need for time course studies in order to better assess the outcome of toxicogenomic studies.

Zidek *et al.* [54] used a bead-based Illumina oligonucleotide microarray containing 550 liver-specific genes to establish a predictive screening system for acute hepatotoxicity *in vivo* (rat). Differential gene expression profiles of well-known hepatotoxic (6) and nonhepatotoxic compounds (6) were generated from liver of Sprague Dawley rats after 6, 24, and 72 h. Based on leave-one-out cross-validation analysis, gene expression profiling allowed the accurate discrimination of all model compounds, 24 h after high-dose treatment. During the regeneration phase, CCl_4, α-naphthyl isothiocyanate (ANIT), and APAP

were predicted to be hepatotoxic, where only these compounds showed histopathological changes. Furthermore, 64 potential marker genes responsible for class prediction were identified, which reflected typical hepatotoxic responses.

These models are able to classify hepatotoxicants into distinct classes, but in reality hepatotoxicity caused by chemicals often consists of a mixture of toxicities.

5.4.4 *In Vitro* Classifiers of Hepatotoxicity

Cha *et al.* [55] have recently identified a classifier for hepatotoxicity prediction, specifically for NSAIDs by analyzing differential gene expression profiles in HepG2 cells. A hepatotoxicity prediction model based on eight positive compounds was built and 77 specific genes identified as being predictive. These genes and pathways, commonly regulated by hepatotoxicants, may be indicative of the early characterization of hepatotoxicity and possibly predictive of later hepatotoxicity onset. Four test compounds, including hepatotoxic and nonhepatotoxic NSAIDs, were used for validating the prediction model and the accuracy was 100%. Even though these results are promising, the gene expression of HepG2 cells does vary greatly from the situation in the liver. The reliability of this small data set and the relevance of the gene signature for the situation in hepatocytes or *in vivo* will have to be confirmed in the future. Chen *et al.* [56] published a study in which a cell-based molecular predictor of hepatotoxicity was developed (based on *in vivo* rat and *in vitro* human data), and tested this signature against data from more than 160 diverse compounds. They were able to predict *in vivo* acute hepatotoxicity from an *in vitro* cell model.

In a recent publication, *in vitro* and molecular techniques were combined to establish a new *in vitro* model for toxicological screening [57]. Similar techniques have been reported before; however, the combination used here was novel and the data set used was more comprehensive than other smaller studies, which addressed only very specific questions [58,59]. The key objective of this study was to determine whether it is possible to distinguish between hepatotoxic and nonhepatotoxic compound-based gene expression data from long-term exposed *in vitro* cultured rat hepatocytes.

5.4.5 Biomarker Identification

Drug-induced hepatotoxicity is a common cause for liver injury and specific biomarkers, which indicate these alterations, are a useful tool to monitor the possible occurrence of liver injury in the clinic. The traditional biomarkers in the clinic are serum alanine transferase (ALT) and aspartate aminotransferase (AST) [60]. However, these biomarkers are neither very sensitive nor specific, since changes may occur only after long dosing periods or may be influenced by other factors. The current biomarkers represent clinical

chemistry endpoints, which display secondary effects of toxicities at late stages. Second, these biomarkers are mostly based on single molecules that can also be altered depending on factors such as age, nutritional state, and alcohol intake. Genomic biomarkers are more reliable and sensitive and combinations of functional and morphological markers would represent an attractive approach for the detection of hepatotoxicity at early stages or in short-term studies. Therefore, toxicogenomics could provide a more sensitive identification and characterization of specific biomarkers, preferably after exposure to low doses and in early stages of drug exposure.

The development of new biomarkers requires an appropriate repository, including the gene expression patterns of multiple, structurally different compounds with similar modes of toxicity as well as nontoxic compounds. Computational algorithms can then be used to train a set for the identification of genes specifically deregulated by the compound class. Validation and the understanding of accuracy, sensitivity, and specificity can be achieved via the generation of new gene expression profiles using compounds not used in the training of the model [61].

Heijne *et al.* [62] investigated bromobenzene-induced hepatic necrosis and revealed altered gene expression levels and metabolites related to the severity of necrosis, thus providing putative novel biomarkers. Therefore, some effects on the expression levels were identified early (6 h) after treatment when no other marker indicated hepatotoxicity. The combined analysis of expression changes and plasma metabolomics showed its ability to detect hepatotoxicant-related changes more sensitively. Since toxicogenomic approaches display a variety of effects, it is essential to understand whether an effect is specific and also associated with toxicity. For the investigation of specificity, Heijne *et al.* [62] compared the gene expression pattern of low-dose bromobenzene with those of model hepatotoxicants, bromobenzene and acetaminophen at high doses. Expression changes in response to low-dose bromobenzene were related to pathways of biotransformation as well as physiological pathways such as fatty acid and cholesterol metabolism, thus showing a substantial overlap with those of high-dose bromobenzene and acetaminophen. This shows that comparing low-dose responses with those of high-dose model hepatotoxicants aids in the identification of new biomarkers.

Biomarkers also display a useful tool to provide better mechanistic understanding. For instance, Ruepp *et al.* [53] investigated genomic classifiers for steatosis derived from expression profiles of rat livers exposed to various model compounds. Three proprietary antidiabetics were identified as steatotic using these markers. Histopathology showed steatotic alterations for two compounds, whereas the third compound was further investigated by various chemical assays, which confirmed the steatotic liability. Furthermore, a repeated dose study in dogs revealed a steatotic potential for the third compound.

This illustrates that these markers are able to identify the adverse potentials of a compound at early stages as well as provide an improved extrapolation of toxic effects across species.

5.5 Toxicogenomic Profiling of Nephrotoxicity

In addition to the liver, the investigation of possible adverse effects of a drug on the kidney is also of major concern during drug development due its variety of functions in homeostasis and the excretion of drug metabolites and endogenous waste products. The kidney performs a wide range of functions, being essential for the regulation of a constant extracellular environment as well as the maintenance of metabolic homeostasis. Therefore, the kidney is one of the routinely assessed organs during preclinical safety assessment.

However, during preclinical development only 7% of drug candidates fail due to nephrotoxicity [63]. In contrast, the incidence of patients developing acute kidney injury (AKI) in intensive care units is between 30 and 50% [64], leading to the assumption that nephrotoxicity may be underestimated during preclinical development. Therefore, the data generated in preclinical phases have to be as detailed as possible in order to support clinicians and minimize the risk of acute kidney injury. AKI is defined as an abrupt and sustained decrease in renal function, resulting in the retention of waste products as well as the deregulation of electrolyte homeostasis. AKI is mostly estimated by a low glomerular filtration rate and a high serum creatinine (SCr) level; however, there are major limitations for the use of these parameters.

The high incidence of AKI in the clinic is also aggravated by the fact that the current biomarkers used for the monitoring of kidney function have some crucial limitations. The classical parameters for the monitoring of patients in clinical studies are primarily based on SCr and blood urea nitrogen (BUN), which remain the gold standard. However, these biomarkers have both poor sensitivity and specificity for the detection of early and acute stages of kidney injury. In addition, SCr and BUN are not applicable for a more specific insight into either the location or the type of injury. Therefore, new biomarkers are needed not only for an earlier detection of AKI but also for the definition of its origin or subtype of AKI.

5.5.1 Toxicogenomic Approaches in Nephrotoxicity

Toxicogenomic approaches can help to solve the two major problems in the clinic:

- Identification of high-risk patients.
- Advanced biomarkers for monitoring kidney function.

In the last decade, many different investigations have been made using toxicogenomics together with classical evaluation in order to discover the risk of patients as well as novel biomarkers, thus providing the clinicians better possibilities to assess renal injury.

5.5.2 Finding Genes that Matter in AKI

Regarding a patient's risk to develop AKI, the understanding of epithelial, vascular, and immune responses makes it likely that genetic variation in the regulatory elements of these responses plays a major role. For instance, genetic variations in inflammatory cytokines such as TNF-α (tumor necrosis factor alpha) and IL-6 (interleukin 6) have been proposed as risk factors for AKI [65,66] due to the role of inflammatory mediators in the pathophysiology of acute kidney injury. However, it is complicated to determine which variations are truly associated with AKI. Different investigations have been made in order to reveal any relationship between genetic polymorphisms and the risk to develop AKI. Isbir *et al.* [67] as well as Chew *et al.* [68] reported that carriers of the apolipoprotein E (ApoE) e4-allele had a decreased risk for the development of AKI after surgery in independent studies.

Several other studies have focused on genes involved in oxidative stress, inflammation, vasomotor stress, and gene–gene interactions, but they did not find any significant consensus, although some trends were revealed. For example, Perianayagam *et al.* [69–71] investigated the association of NADPH (nicotinamide adenine dinucleotide phosphatenicotinamide adenine dinucleotide phosphate) oxidase and catalase polymorphisms with susceptibility to AKI. The carriers of the T-allele of the p22phox subunit of NADPH oxidase showed the trend to have a higher risk for chronic dialysis, although there was no significance. In contrast, Isbir *et al.* [67] focused on whether a relationship between angiotensin-converting enzyme (ACE) I/D polymorphism and AKI exists. They showed that carriers of the ACE D-allele have an increased risk for developing AKI. However, no association of this polymorphism was found in other studies [72].

The possible role of both inflammatory and anti-inflammatory genes was investigated by Jaber *et al.* [65], who showed an increased risk for high producers of TNF-α, which is a proinflammatory cytokine. In contrast, this study also revealed a decreased risk for individuals who produce high levels of interleukin 10 (IL-10), which suppresses the acute immune response. Other studies could not find any clear association between these mediators and susceptibility to AKI [72,73]. This is also an example of how gene–gene interactions may play a role, since the combination of low TNF-α and high IL-10 phenotypes would possibly have the lowest risk. Polymorphisms of IL-10 were also studied by Wattanathum *et al.* [74], who could separate three polymorphisms that are associated with a higher susceptibility. This has led to the hypothesis that the lower IL-10 production by these individuals is associated with a higher risk due to the anti-inflammatory properties of IL-10.

However, all these studies could not reveal that any single polymorphism can be conclusively associated with an increased risk for developing AKI.

5.5.3 Searching for New Biomarkers of Kidney Injury

Next to genomics-based investigations focusing on high-risk patients, toxicogenomic approaches have specialized in the discovery of novel biomarkers for the early and reliable detection of AKI. Ideally, new biomarkers should have the following properties [75]:

- Noninvasive or minimally invasive sample collection (blood, urine).
- Higher sensitivity at earlier stages (e.g., absence in healthy kidney).
- More specific identification of damaged area (prerenal, intrarenal, postrenal).
- Easy analysis, and rapid and reliable measurement with standard equipment.

For this urgent need, several initiatives and consortia have addressed this issue and worked on the identification of novel biomarkers using toxicogenomic approaches. The Predictive Safety Testing Consortium (PSTC) was the first group to formally submit and have qualified a novel set of biomarkers by the regulatory agencies (in this case for acute rat nephrotoxicity) [76]. Seven renal safety biomarkers were qualified for limited use in nonclinical and clinical drug development to help guide safety assessments. Of specific significance were KIM-1 (kidney injury molecule-1), clusterin, and albumin, which were shown to individually outperform the old gold standard blood urea nitrogen and serum creatinine assays as early diagnostic biomarkers of drug-induced acute tubular necrosis. In addition, urinary TFF3 (trefoil factor 3) can add information. Total urinary protein, cystatin C, and β_2-microglobulin are better early diagnostic biomarkers of acute drug-induced glomerular alterations or damage resulting in impairment of kidney tubular reabsorption.

The ILSI/HESI (International Life Sciences Institute's Health and Environmental Sciences Institute) Nephrotoxicity Working Group was founded in 1999 with the aim to identify renal-specific biomarkers at the transcriptomic and proteomic levels, including microarray analysis. Initially, gene expression data of three rat toxicity studies of known model compounds (puromycin, gentamicin, and cisplatin) were generated and compared with classical histopathological and clinical chemical parameters [77,78]. As a result, many genes showed altered expression in treated animals and were initially considered as being potentially novel biomarkers. Subsequently, an initial validation process was initialized in order to elucidate valuable associations between biomarker and injury. Subsequently, the group investigated 10 compounds, out of which 8 are well-known nephrotoxicants and 2 hepatotoxic compounds served as negative controls. In this study, 23 potential new urinary biomarkers and many transcriptional biomarkers were identified. The validation has led to seven biomarkers (an eight was later qualified), which have been qualified by regulatory authorities (FDA, European Medicines Agency (EMA), and Product Development and Management Association (PDMA)) for the supportive use in short-term rat toxicity studies up to 14 days.

KIM-1 is a transmembrane glycoprotein, which is not detectable in healthy kidney tissue or urine, but is highly expressed in human and rodent proximal tubule cells after ischemic or toxic injury [79]. It has been reported that KIM-1 is upregulated 24–48 h postischemia in rat kidneys and Han *et al.* [80] detected that KIM-1 is measurable in urine of patients suffering from acute tubular necrosis, thus facilitating the early diagnosis of proximal tubule injury. The major advantage of KIM-1 as a novel biomarker is its noninvasive detection, since the ectodomain is spliced off and thus present in the urine. Bonventre [81] defined KIM-1 as an attractive biomarker due to various reasons:

- Absence in healthy kidney.
- Significant upregulation in damaged tubules.
- Persistence until the cell has recovered.
- Noninvasive detection of temperature stability of ectodomain for *ex vivo* analysis.

Cystatin C is a cysteine protease inhibitor that is synthesized by all nucleated cells and is freely filtered by the glomerulus and completely reabsorbed by proximal tubule cells. In contrast to serum creatinine, the levels of cystatin C are not affected by age, gender, race, or muscle mass and it was found out that this emerging biomarker is able to detect acute kidney injury almost 2 days earlier compared with SCr in high-risk patients [82]. Koyner and Parikh [83] also reported that urinary cystatin C seems to be a promising biomarker for AKI following cardiac surgery by being detectable within 6 h after surgery.

NGAL (neutrophil gelatinase-associated lipocalin) is a protein of the lipocalin family, mainly expressed by neutrophils and epithelia including proximal tubule cells. Supavekin *et al.* [84] identified NGAL as one of the most rapidly upregulated genes in the postischemic mouse kidney, which is also extensively detectable in mouse urine after cisplatin treatment [85]. NGAL is a useful and early predictor of AKI, also having prognostic value for clinical endpoints [86]. However, limitations of this biomarker have also been identified and extrarenal generation may occur due to systemic stress, thus increasing the urine level of NGAL in the absence of AKI.

Vinken *et al.* [87] investigated the characteristics of traditional (SCr and BUN) and novel (KIM-1, NGAL, clusterin, etc.) biomarkers in a 14-day study in male Sprague Dawley rats treated with 1 mg cisplatin/kg/day. Urinalysis, blood analysis, and histopathology as well as immunohistochemistry were compared. The traditional biomarkers did not occur before day 5 of treatment, whereas urinary clusterin as well as tissue KIM-1 increased significantly after 1 day of dosing before any histopathological changes occurred and urinary KIM-1 was detectable at day 3. Together with traditional histopathology, KIM-1 and clusterin could be defined as early, specific and sensitive biomarkers for acute kidney injury in rats. However, in this study, NGAL did not show this early significant increase before day 5 and levels of significance were lower than those of SCr and BUN.

Identification of novel early markers of calcineurin inhibitor (CNI) nephrotoxicity depends on understanding of the early biological response of stressed kidney tissue. Recently, transcriptomic analysis of *in vitro* models of CNI nephrotoxicity has identified two original molecular mechanisms that may play a role in early CNI-induced nephrotoxicity and serve as early recognition markers: the epithelial-to-mesenchymal transition (EMT) and endoplasmic reticulum (ER) stress [88].

A limited number of papers have been published describing *in vitro* efforts to combine omics analysis and understanding nephrotoxicity. Wilmes *et al.* [89] investigated the added benefit of integrating transcriptomic, proteomic, and metabolomic data streams for the application to predictive *in vitro* toxicology and safety assessment, in relation to the nephrotoxicity of cyclosporin A (CSA). Cultured human renal proximal tubule cells were treated with the immunosuppressive and nephrotoxic compound CsA for 1, 3, and 14 days in a 24 h repeated dose testing regime. The integrated analysis of these data showed that 15 μM CsA, but not 5 μM (therapeutic level), induced mitochondrial disturbances and activated the Nrf2-mediated oxidative stress pathway and the unfolded protein response. In addition, both the Nrf2 and PERK/ATF4 signatures were clearly present in all omics data methods. These authors concluded that their study was the first to really show that the combination of different omics techniques has great potential for deciphering and understanding the mechanisms of compound-induced cell stress.

5.6 Limitations of Toxicogenomics

The applications of toxicogenomic approaches display many benefits, not only in drug development but also for all aspects in biological and medical research. However, toxicogenomics has some limitations as well, since it is not applicable for all questions. In addition, the half-life of mRNA can be quite short and thus some features may be missed. The following briefly describes two important limitations.

5.6.1 Idiosyncrasies

Idiosyncratic toxicity is defined as an unpredictable, characteristic adverse reaction of an individual, usually occurring at very low frequencies. These unpredictable effects pose a major issue in all phases of clinical drug development and especially during postmarketing phases. Idiosyncratic reactions occur only in a small proportion of individuals exposed to therapeutic doses, usually ranging from 0.1 to 0.01%, thus not occurring until a vast number of patients have been exposed to the drug (clinical phase III, postmarketing).

This explains why idiosyncrasies are the most common reason for postmarketing warnings, so-called black-box warnings, as well as for withdrawals and about 10% of acute liver failures are due to idiosyncratic hepatotoxicity. In general, these effects are not related to the pharmacological target of a drug and the underlying mechanisms are not fully understood. It is assumed that both genetic and environmental factors account for these reactions; in addition, it is assumed that the immune system plays an important role in these rare adverse effects. The application of toxicogenomics in this field is quite difficult and has limited utility. The revelation of genetic risk factors does not elucidate the functional interaction among individuals and also the study of environmental factors is not possible.

5.6.2 Epigenetics

The discipline of epigenetics concerns all interactions between the genome of an organism and the environment. This includes all processes that lead to different gene activities and thus a different outcome (phenotype). However, these alterations do not occur at the base pair level, meaning that the DNA sequence itself is not affected. The key mechanisms of epigenetic change lie within DNA methylation, histone modifications, and miRNAs, which lead to altered activity of sequence areas. For instance, in toxicology, epigenetic alterations may act as early biomarkers for genotoxic and nongenotoxic carcinogens. Furthermore, the gene regulation of transporters, phase I and phase II metabolizing enzymes, and transcription factors is epigenetically encoded, thus emphasizing the role of environmental influences and the importance in potentially differing drug responses. The field of epigenetics therefore addresses how tissue-specific gene regulations interact with environmental stimuli, having the potential to explain the different and individual responses to drug treatment. Epigenomic approaches help to identify these modifications, for example, by labeling methylations gaining knowledge of how an alteration may be related to a specific phenotype.

However, although investigations in this field may help complete the complex puzzle of interactions between an individual and a xenobiotic, it has to be kept in mind that these investigations display a snapshot in time and thoughts such as individual profiling of every patient remain unfeasible at present due to the complexity of these processes and have limited applicability in the field.

5.7 Conclusions

Toxicogenomics is a relatively new scientific field, with several years of technology optimization and subsequent gaining of experience by multiple researchers. Overall, toxicogenomics helps the toxicologist to study xenobiotic-mediated gene expression changes, especially at subtoxic, therapeutic doses. The focus

thus lies on the revelation of changes in the gene expression levels prior to pathological changes. These applications can enable a better understanding of mechanisms leading to toxicities and can help to better predict possible adverse effects of unknown compounds. Furthermore, toxicogenomic research helps to define mechanism-based, tissue-specific biomarkers and compound classification based on expression signatures. Therefore, the overall aim is to systematically understand a biological response to a toxicant prior to its pathological outcome. This can be summarized as making drug development more efficient by supporting the classical "gold standard" methods and also circumventing their known limitations.

In addition to the strengths and benefits, toxicogenomic applications also have limitations and potential pitfalls. One of the major challenges lies in the proper incorporation and implementation of toxicogenomic technologies. The dimensions of such extraordinarily large amounts of data display a new dimension of possible errors, including proper strategies of probe annotations and data comparison across different platforms. The proper interpretation is also a challenge since not all significant expression changes are physiologically relevant and the expertise of the scientists is essential for a full and appropriate data interpretation.

References

1 Chin, K.V. and Kong, A.N. (2002) Application of DNA microarrays in pharmacogenomics and toxicogenomics. *Pharm. Res.*, **19**, 1773–1778.

2 Nuwaysir, E.F., Bittner, M. *et al.* (1999) Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.*, **24**, 153–159.

3 Lord, P.G. and Nie, A. (2006) Application of genomics in preclinical drug safety evaluation. *Basic Clin. Pharmacol. Toxicol.*, **98**, 537–546.

4 Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.*, **162**, 156–159.

5 Banerjee, M. and Giri, A.K. (2009) An overview with special reference to genetic and genomic approaches to the identification of toxic effects, in *General Applied and Systems Toxicology*, John Wiley & Sons, Ltd, Chichester.

6 Binder, H., Krohn, K., and Burden, C.J. (2010) Washing scaling of GeneChip microarray expression. *BMC Bioinformatics*, **11**, 291.

7 Horn, T., Chang, C.A., and Urdea, M.S. (1997) Chemical synthesis and characterization of branched oligodeoxyribonucleotides (bDNA) for use as signal amplifiers in nucleic acid quantification assays. *Nucleic Acids Res.*, **25**, 4842–4849.

8 Tsongalis, G.J. (2006) Branched DNA technology in molecular diagnostics. *Am. J. Clin. Pathol.*, **126**, 448–453.

9 Butte, A. (2002) The use and analysis of microarray data. *Nat. Rev. Drug Discov.*, **1**, 951–960.

10 MAQC Consortium (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurement. *Nat. Biotechnol.*, **24** (9), 1151–1161.

11 MAQC Consortium (2010) MAQC-II: analyze that! *Nat. Biotechnol.*, **28** (8), 761.

12 Ulrich, R. and Friends, S.H. (2002) Toxicogenomics and drug discovery: will new technologies help us produce better drugs? *Nat. Rev. Drug Discov.*, **1**, 84–88.

13 Matsumoto, T. and Kawakami, M. (1982) The unit-concept of hepatic parenchyma – a re-examination based on

angioarchitectural studies. *Acta Pathol. Jpn.*, **32** (Suppl. 2), 285–314.
14 Shackel, N.A., Gorrell, M.D. *et al.* (2002) Gene array analysis and the liver. *Hepatology*, **36**, 1313–1325.
15 Kogure, K. and Ishizaki, M. (1999) A comparative study of the anatomy of rat and human livers. *J. Hepatobiliary Pancreat. Surg.*, **6**, 171–175.
16 Burt, A.D., Day, C.P. *et al.* (2002) Pathophysiology of the liver, in *Pathology of the Liver*, Churchill Livingstone, New York, pp. 67–105.
17 Stuart, E.T. and Wheatley, A.M. (1995) Redistribution of portal venous but not hepatic arterial flow is induced by hepatic nerve stimulation in the perfused rat liver. *Arch. Physiol. Biochem.*, **103**, 99–108.
18 Thein, E., Becker, M. *et al.* (2003) Direct assessment and distribution of regional portal blood flow in the pig by means of fluorescent microspheres. *J. Appl. Physiol.*, **95**, 1808–1816.
19 Daniel, G.B., DeNovo, R.C. *et al.* (2004) Portal streamlining as a cause of nonuniform hepatic distribution of sodium pertechnetate during per-rectal portal scintigraphy in the dog. *Vet. Radiol. Ultrasound*, **45**, 78–84.
20 Faa, G., Sciot, R. *et al.* (1994) Iron concentration and distribution in the newborn liver. *Liver*, **14**, 193–199.
21 Faa, G., Nurchi, V. *et al.* (1995) Uneven hepatic copper distribution in Wilson's disease. *J. Hepatol.*, **22**, 303–308.
22 Ambu, R., Crisponi, G. *et al.* (1995) Uneven hepatic iron and phosphorus distribution in beta-thalassemia. *J. Hepatol.*, **23**, 544–549.
23 Matsuzaki, S., Onda, M. *et al.* (1997) Hepatic lobar differences in progression of chronic liver disease: correlation of asialoglycoprotein scintigraphy and hepatic functional reserve. *Hepatology*, **25**, 828–832.
24 Richardson, F.C. and Boucheron, J.A. (1986) Biochemical and morphologic studies of heterogeneous lobe responses in hepatocarcinogenesis. *Carcinogenesis*, **7**, 247–251.
25 Heinloth, A.M., Irwin, R.D. *et al.* (2004) Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol. Sci.*, **80**, 193–202.
26 Malarkey, D.E., Johnson, K. *et al.* (2005) New insights into functional aspects of liver morphology. *Toxicol. Pathol.*, **33**, 31–37.
27 Malarkey, D.E., Parker, J.S. *et al.* (2005) Microarray data analysis of mouse neoplasia. *Toxicol. Pathol.*, **33**, 127–135.
28 Teutsch, H.F., Schuerfeld, D., and Groezinger, E. (1999) Three-dimensional reconstruction of parenchymal units in the liver of the rat. *Hepatology*, **29**, 494–505.
29 Jungermann, K. and Katz, N. (1989) Functional specialization of different hepatocyte populations. *Physiol. Rev.*, **69**, 708–764.
30 Gebhardt, D.O. and Souverijn, J.H. (1992) Heterogeneity of lipoprotein A-I. *Clin. Chem.*, **38**, 2345.
31 Pennie, W.D. (2000) Use of cDNA microarrays to probe and understand the toxicological consequences of altered gene expression. *Toxicol. Lett.*, **112** (1–3), 473–477.
32 Burchiel, S.W., Knall, C.M. *et al.* (2001) Analysis of genetic and epigenetic mechanisms of toxicity. Potential roles of toxicogenomics and proteomics in toxicology. *Toxicol. Sci.*, **59**, 193–195.
33 Hamadeh, H.K., Bushel, P. *et al.* (2001) Discovery in toxicology: mediation by gene expression array technology. *J. Biochem. Mol. Toxicol.*, **15** (5), 231–242.
34 Hamadeh, H.K., Bushel, P.R. *et al.* (2002) Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.*, **67**, 232–240.
35 Bulera, S.J., Eddy, S.M. *et al.* (2001) RNA expression in the early characterization of hepatotoxicants in Wistar rats by high-density DNA microarrays. *Hepatology*, **33**, 1239–1258.
36 Hamadeh, H.K., Bushel, P.R. *et al.* (2002) Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.*, **67**, 219–231.
37 Hamadeh, H.K., Amin, R.P. *et al.* (2002) An overview of toxicogenomics. *Curr. Issues Mol. Biol.*, **4**, 45–56.
38 Burczynski, M.E., McMillian, M. *et al.* (2000) Toxicogenomics-based discrimination of toxic mechanism in

HepG2 human hepatoma cells. *Toxicol. Sci.*, **58**, 399–415.

39 Ellinger-Ziegelbauer, H., Stuart, B. *et al.* (2004) Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol. Sci.*, **77**, 19–34.

40 McMillian, M., Nie, A.Y. *et al.* (2005) Drug-induced oxidative stress in rat liver from a toxicogenomics perspective. *Toxicol. Appl. Pharmacol.*, **207** (Suppl. 2), 171–178.

41 Kong, A.N., Owuor, E. *et al.* (2001) Induction of xenobiotic enzymes by the MAP kinase pathway and the antioxidant or electrophile response element (ARE/EpRE). *Drug Metab. Rev.*, **33**, 255–271.

42 Boess, F., Durr, E. *et al.* (2007) An *in vitro* study on 5-HT6 receptor antagonist induced hepatotoxicity based on biochemical assays and toxicogenomics. *Toxicol. In Vitro*, **21**, 1276–1286.

43 Lum, P.Y., He, Y.D. *et al.* (2007) Gene expression profiling of rat liver reveals a mechanistic basis for ritonavir-induced hyperlipidemia. *Genomics*, **90**, 464–473.

44 Jolly, R.A., Ciurlionis, R. *et al.* (2004) Microvesicular steatosis induced by a short chain fatty acid: effects on mitochondrial function and correlation with gene expression. *Toxicol. Pathol.*, **32** (Suppl. 2), 19–25.

45 Liguori, M.J., Anderson, M.G. *et al.* (2005) Microarray analysis in human hepatocytes suggests a mechanism for hepatotoxicity induced by trovafloxacin. *Hepatology*, **41**, 177–186.

46 Dai, X., He, Y.D. *et al.* (2006) Development of an approach for *ab initio* estimation of compound-induced liver injury based on global gene transcription profiles. *Genome Inform.*, **17** (2), 77–88.

47 Blomme, E.A., Yang, Y., and Waring, J.F. (2009) Use of toxicogenomics to understand mechanisms of drug-induced hepatotoxicity during drug discovery and development. *Toxicol. Lett.*, **186**, 22–31.

48 Jaeschke, H. (2002) Redox considerations in hepatic injury and inflammation. *Antioxid. Redox Signal.*, **4**, 699–700.

49 Pessayre, D., Mansouri, A. *et al.* (1999) Hepatotoxicity due to mitochondrial dysfunction. *Cell Biol. Toxicol.*, **15**, 367–373.

50 Suter, L., Haiker, M. *et al.* (2003) Effect of two 5-HT6 receptor antagonists on the rat liver: a molecular approach. *Pharmacogenomics J.*, **3**, 320–334.

51 McMillian, M., Nie, A.Y. *et al.* (2004) Inverse gene expression patterns for macrophage activating hepatotoxicants and peroxisome proliferators in rat liver. *Biochem. Pharmacol.*, **67**, 2141–2165.

52 Waring, J.F., Jolly, R.A. *et al.* (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.*, **175**, 28–42.

53 Ruepp, S., Boess, F. *et al.* (2005) Assessment of hepatotoxic liabilities by transcript profiling. *Toxicol. Appl. Pharmacol.*, **207** (Suppl. 2), 161–170.

54 Zidek, N., Hellmann, J. *et al.* (2007) Acute hepatotoxicity: a predictive model based on focused Illumina microarrays. *Toxicol. Sci.*, **99**, 289–302.

55 Cha, R.H., Lee, C.S. *et al.* (2010) Clinical usefulness of serum cystatin C and the pertinent estimation of glomerular filtration rate based on cystatin C. *Nephrology*, **15**, 768–776.

56 Chen, M. *et al.* (2011) FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov. Today*, **16**, 697–703.

57 Hrach, J., Mueller, S.O., and Hewitt, P. (2011) Development of an *in vitro* liver toxicity prediction model based on long-term primary rat hepatocyte culture. *Toxicol. Lett.*, **206**, 189–196.

58 Boess, F., Kamber, M. *et al.* (2003) Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the *in vivo* liver gene expression in rats: possible implications for toxicogenomics in use of *in vitro* systems. *Toxicol. Sci.*, **73**, 386–402.

59 Braeuning, A., Ittrich, C., *et al.* (2006) Differential gene expression in periportal and perivenous mouse hepatocytes. *FEBS J.*, **273** (22), 5051–5061.

60 Boone, L., Meyer, D. *et al.* (2005) Selection and interpretation of clinical pathology indicators of hepatic injury in preclinical studies. *Vet. Clin. Pathol.*, **34**, 182–188.

61 Yang, H.H., Hu, Y. *et al.* (2004) A computational approach to measuring

coherence of gene expression in pathways. *Genomics*, **84**, 211–217.
62 Heijne, W.H., Jonker, D. *et al.* (2005) Toxicogenomic analysis of gene expression changes in rat liver after a 28-day oral benzene exposure. *Mutat. Res.*, **575**, 85–101.
63 Frost & Sullivan (2007) *Drug Discovery and Development* (ed. H.P. Ranf), Elsevier/Churchill Livingstone.
64 Devarajan, P. (2005) Cellular and molecular derangements in acute tubular necrosis. *Curr. Opin. Pediatr.*, **17**, 193–199.
65 Jaber, B.L. and Liangos, O. (2004) Polymorphism of immunomodulatory cytokine genes: implications in acute renal failure. *Blood Purif.*, **22**, 101–111.
66 Jaber, B.L., Rao, M. *et al.* (2004) Cytokine gene promoter polymorphisms and mortality in acute renal failure. *Cytokine*, **25**, 212–219.
67 Isbir, S.C., Tekeli, A. *et al.* (2007) Genetic polymorphisms contribute to acute kidney injury after coronary artery bypass grafting. *Heart Surg. Forum*, **10**, E439–444.
68 Chew, S.T. *et al.* (2000) Preliminary report on the association of apolipoprotein E polymorphisms, with postoperative peak serum creatinine concentrations in cardiac surgical patients. *Anesthesiology*, **93**, 325–331.
69 Perianayagam, M.C., Liangos, O. *et al.* (2006) NADPH oxidase p22phox and catalase gene variants are associated with biomarkers of oxidative stress and adverse outcomes in acute renal failure. *J. Am. Soc. Nephrol.*, **18**, 255–263.
70 Perianayagam, M.C., Seabra, V.F. *et al.* (2009) Serum cystatin C for prediction of dialysis requirement or death in acute kidney injury: a comparative study. *Am. J. Kidney Dis.*, **54**, 1025–1033.
71 Perianayagam, M.C. and Tighiouart, H. (2012) Polymorphisms in the myeloperoxidase gene locus are associated with acute kidney injury-related outcomes. *Kidney Int.*, **82**, 909–919.
72 Stafford-Smith, M. (2013) Can "earlier biomarkers" help early biomarkers predict acute kidney injury? *Crit. Care Med.*, **41**, 914–915.
73 Treszl, A., Toth-Heyn, P. *et al.* (2002) Interleukin genetic variants and the risk of renal failure in infants with infection. *Pediatr. Nephrol.*, **17**, 713–717.
74 Wattanathum, A., Manocha, S. *et al.* (2005) Interleukin-10 haplotype associated with increased mortality in critically ill patients with sepsis from pneumonia but not in patients with extrapulmonary sepsis. *Chest*, **128**, 1690–1698.
75 Nguyen, M.T. and Devarajan, P. (2008) Biomarkers for the early detection of acute kidney injury. *Pediatr. Nephrol.*, **23**, 2151–2157.
76 Dieterle, F., Sistare, F., Goodsiad, F. *et al.* (2010) Renal biomarker qualification submission: a dialog between the FDA–EMEA and Predictive Safety Testing Consortium. *Nat. Biotechnol.*, **28** (5), 455–462.
77 Amin, R.P., Vickers, A.E. *et al.* (2004) Identification of putative gene based markers of renal toxicity. *Environ. Health Perspect.*, **112**, 465–479.
78 Thompson, K.L., Afshari, C.A. *et al.* (2004) Identification of platform-independent gene expression markers of cisplatin nephrotoxicity. *Environ. Health Perspect.*, **112**, 488–494.
79 Ichimura, T., Bonventre, J.V. *et al.* (1998) Kidney injury molecule-1 (KIM-1), a putative epithelial cell adhesion molecule containing a novel immunoglobulin domain, is upregulated in renal cells after injury. *J. Biol. Chem.*, **273**, 4135–4142.
80 Han, W.K., Bailly, V. *et al.* (2002) Kidney injury molecule-1 (KIM-1): a novel biomarker for human renal proximal tubule injury. *Kidney Int.*, **62**, 237–244.
81 Bonventre, J.V. (2009) Kidney injury molecule-1 (KIM-1): a urinary biomarker and much more. *Nephrol. Dial Transplant*, **24**, 3265–3268.
82 Dharnidharka, V.R., Kwon, C., and Stevens, G. (2002) Serum cystatin C is superior to serum creatinine as a marker of kidney function: a meta-analysis. *Am. J. Kidney Dis.*, **40**, 221–226.
83 Koyner, J.L. and Parikh, C.R. (2013) Clinical utility of biomarkers of AKI in cardiac surgery and critical illness. *Clin. J. Am. Soc. Nephrol.*, **8**, 1034–1042.

84 Supavekin, S., Zhang, W. *et al.* (2003) Differential gene expression following early renal ischemia/reperfusion. *Kidney Int.*, **63**, 1714–1724.
85 Mishra, J., Ma, Q. *et al.* (2003) Identification of neutrophil gelatinase-associated lipocalin as a novel early urinary biomarker for ischemic renal injury. *J. Am. Soc. Nephrol.*, **14** (10), 2534–2543.
86 Haase, M., Bellomo, R. *et al.* (2009) Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: a systematic review and meta-analysis. *Am. J. Kidney Dis.*, **54**, 1012–1024.
87 Vinken, P., Starckx, S. *et al.* (2012) Tissue Kim-1 and urinary clusterin as early indicators of cisplatin-induced acute kidney injury in rats. *Toxicol. Pathol.*, **40**, 1049–1062.
88 Pallet, N. (2011) New insights on stress-induced epithelial phenotypic changes. *Nephrol. Dial. Transplant.*, **27** (2), 483–485.
89 Wilmes, A., Leonard, M.O., and Jennings, P. (2013) Nrf2 inducibility of aldo-keto reductases. *Toxicol. Lett.*, **221**, 39.

6
Use of Toxicogenomics for Mechanistic Characterization of Hepatocarcinogens in Shorter Term Studies

Heidrun Ellinger-Ziegelbauer

6.1
Introduction

6.1.1
Rodent Carcinogenicity Testing

Carcinogenicity testing of drugs in development, but also of environmental and other chemicals, currently still relies on lifetime studies in rats and mice, often abbreviated as 2-year rodent bioassays. For over 50 years now, this assay has been the regulatory standard for prediction of potential human cancer hazard [1]. The recommended study design, which has been refined and detailed over time, is rather similar for all chemicals that require regulatory testing, for example, environmental, agricultural, industrial, or pharmaceutical chemicals. These study designs are laid down in "International Conference on Harmonisation" (ICH) guidelines S1A, S1B, and S1C for pharmaceuticals [2–4] and in "Organization for Economic Co-operation and Development" (OECD) guidelines in general [5]. Normally, two species must be studied, with at least one study, in general the rat, in a long-term (2-year) setting. The second study may be a 2-year bioassay in mice, but may also be a 6-month study in a transgenic mouse model. For the 2-year rat bioassay, sufficiently large sample sizes must be selected to ensure 25 live animals per sex per dose level at the end of the study after 2 years (summarized in Ref. [6]). Three dose levels and at least one control should be included. Specific doses need to be selected in prior subchronic, that is, 13-week, studies. Criteria for dose determination are described in ICH S1C(R2) [4].

Overall, the requirements for the number of animals, time, and budget are considerable. For these and other reasons, including animal welfare and scientific aspects, recent initiatives have reevaluated the wealth of data now available from such 2-year bioassays in comparison with data from shorter term studies and other knowledge with respect to carcinogenic risk assessment of human pharmaceuticals. One of the first evaluations was performed by the Pharmaceutical Research and Manufacturing Association (PhRMA) suggesting the so-called Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathologic

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

Risk Factors for Carcinogenicity in the Rat (NEGCARC) approach [6–8]. It is suggested in this approach that only little value is gained from a 2-year rat bioassay for compounds that lack (1) histopathological risk factors for rat neoplasia in chronic, mostly 6-month, toxicology studies, (2) evidence of hormonal perturbation, and (3) positive genetic toxicology results [8]. Depending on the exact details, 82–88% of the compounds predicted as noncarcinogens were true negatives, that is, did not show any tumor induction after 2 years in the rat. For the 18% false-negative compounds, it was concluded that the 2-year study findings were of rather questionable human relevance [8]. Overall, this would have saved 2-year studies for 40% of the drugs.

These insights, including potential shortcomings of the NEGCARC approach, have been taken up by regulatory authorities, who have made their own assessments (summarized in Ref. [9]). Finally, the ISH steering team published a concept paper in 2012 proposing a change to the current S1 guidelines [10], and the ISH S1 expert working group released a regulatory notice document proposing changes to rodent carcinogenicity testing of pharmaceuticals in August 2013 [11], based on the retrospective analysis of various data sets (PhRMA, Food and Drug Administration (FDA), Japanese Pharmaceutical Manufacturers Association (JPMA), and European Medicines Agency (EMA)). This document suggests a weight of evidence assessment to predict the outcome and value of a 2-year rat bioassay based on consideration of several factors, including, among others, genetic toxicology results, target pharmacology, data from chronic rodent and nonrodent studies, knowledge of potential hormonal perturbation, and results from special studies and endpoints. The goal is to reduce the need for the 2-year rodent carcinogenicity study, through a waiver request by the drug-developing company in cases where the 2-year study would not add real value for human cancer risk assessment. The ICH regulatory document proposes to evaluate carcinogenic assessment documents (CADs) for a period of about 2 years during which sponsors would submit CADs to drug regulatory authorities (DRAs) based on weight of evidence factors. These would be assembled before completion of the 2-year bioassay, which would be reported after review of the CAD by the DRA. This prospective evaluation procedure would simulate the real situation and thus would enable the DRA to gain experience and critically assess the currently proposed changes to ICH S1.

The ICH proposal for changes to the current carcinogenicity assessment suggests weight of evidence factors that include results from special studies and endpoints. These may comprise new biomarkers or results from emerging technologies, and mechanistic insight derived from alternative tests. Omics profiling could contribute such insight and allow better evaluation of rodent findings with respect to human relevance. In this respect, the following will review several aspects of the use of toxicogenomics for the characterization of hepatocarcinogens in shorter term studies, including mechanistic and predictive analyses, the potential of databases recently made available for public access [12], transcriptional benchmark dose modeling [13], and the experience gained from several years of especially transcriptomic analyses in studies with rodent carcinogens.

6.1.2 Classes of Carcinogens

Carcinogens are generally divided into two major categories, genotoxic and nongenotoxic, based on the overall mode of actions, which also has implications for regulatory aspects [14,15]. Genotoxic carcinogens (GCs) interact directly with DNA, mostly after being metabolized in to the ultimate carcinogen, are therefore mutagenic when the mutation is fixed by at least two rounds of DNA replication, and are thus involved in the initiation of tumor formation. They are usually detected with *in vitro* and *in vivo* genotoxicity assays. Their dose response is assumed to be linear; thus, a threshold dose is supposed to be absent. Nongenotoxic carcinogens (NGCs) do not directly interact with DNA and act via multiple mechanisms, including nuclear receptor activation, immune suppression, hormonal perturbation, and induction of cellular damage leading to regenerative hyperplastic effects. They likely promote tumor formation by, for example, supporting clonal expansion of initiated cells and induction of cell proliferation or inhibition of apoptosis. In general, they need to be present for prolonged times and at rather high doses to exert their promoting effect, show nonlinear dose responses, and thus are associated with a threshold dose [14,15]. The gene expression studies summarized in the following sections support this overall classification, but also indicate that there are cases in between, and that some compounds show aspects of both mechanisms.

6.2 Toxicogenomics

Toxicogenomics in a broader sense represents the application of "omics" techniques to toxicological research and risk assessment [16]. In the following, examples for mechanistic and predictive toxicogenomic studies used to characterize and classify rodent hepatocarcinogens in shorter term studies will be described, including their limitations. Recently, several databases containing transcriptomic profiles after treatment of rats with different classes of carcinogens have been released into the public, opening up new opportunities for various analyses. Finally, a rather pragmatic approach using benchmark dose (BMD) modeling applied to gene expression data to derive mechanistically defined threshold doses for rodent carcinogens will be described. This approach was developed by Thomas and coworkers [17] and indicates that the strength of toxicogenomics lies in the area of mechanism investigations.

6.2.1 Mechanistic Toxicogenomic Analysis after Short-Term Treatment with Rodent Hepatocarcinogens

Most studies investigating toxicogenomic profiles after treatment with carcinogens or other compounds employed transcriptomics as a profiling method.

Initially, the mRNA expression changes induced by GCs were characterized. Although GCs are in general recognized with genotoxicity testing, and their molecular mechanisms of action are in principle defined, their analysis allowed to test the profiling method including the interpretations that can be drawn, and in fact defined a relatively clear profile of a DNA damage response that can be used to clearly distinguish GCs from NGCs, or reveal partial genotoxic-like responses of NGCs, which are still different from directly DNA-reactive genotoxins.

In one study, molecular changes induced by *N*-nitrosomorpholine (NNM), a classical rat genotoxic hepatocarcinogen, were investigated during a 7-week treatment period followed by 43 weeks without exposure, that is, a total observation period of 50 weeks with eight analysis time points [18,19]. Gene expression profiles, which were measured with Affymetrix rat RG-U34A microarrays, revealed an early DNA damage response characterized by clear upregulation of target genes of the p53 tumor suppressor, accompanied by increased expression of mRNAs encoding anti-apoptotic proteins and proteins such as neuregulin and cyclin D1 that altogether may result in cell survival (see Figure 3 and Table 1 in Ref. [18]). Necrotic changes that were also induced by NNM during the treatment period were likely reflected in the stress response genes found upregulated during treatment. Interestingly, at the two latest time points (30 and 50 weeks) more than 20 weeks after treatment had stopped, mRNA encoding glutathione *S*-transferase P1 (GSTP1) and the H19 imprinted maternally expressed transcript were found to be increased relative to time-matched controls. In the case of GSTP1, which is a well-known marker of preneoplastic and neoplastic lesions [20], this correlated with protein expression and the area fraction of GSTP-positive foci of altered hepatocytes (see Figure 8 in Ref. [18]). The H19 genomic locus is an imprinted region whose expression from the parental alleles is regulated by differential DNA methylation [21]. Increased expression of this gene later on after treatment with NNM may thus indicate epigenetic changes of the DNA methylation status.

A shorter term study investigating liver mRNA profiles in rats treated with four genotoxic rodent hepatocarcinogens, 2-nitrofluorene (2-NF), dimethylnitrosamine (DMN), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), and aflatoxin B1 (AB1), extracted a rather similar DNA damage response in the rodent liver indicated (Figure 6.1) by upregulation of several characteristic p53 target genes, including the cyclin-dependent kinase inhibitor 1A (Cdkn1a, also known as p21), cyclin G1 (Ccng1), ubiquitin E3 ligase Mdm2, pleckstrin homology-like domain family A member 3 (Phlda3), and B-cell translocation gene 2 (BGT2). Furthermore, mRNA encoding O^6-methylguanine-DNA methyltransferase (Mgmt) was upregulated by essentially all GCs examined [22–24]. Although Mgmt is known to be induced by genotoxic stress *in vivo* and *in vitro*, it does not appear to be a classic p53 target gene. Mgmt expression was found to be regulated by transcription factors also activated by DNA damage, such as nuclear factor κB (NF-κB), with p53 still having a modifying influence [25]. A very similar DNA damage response pattern at the gene expression level was

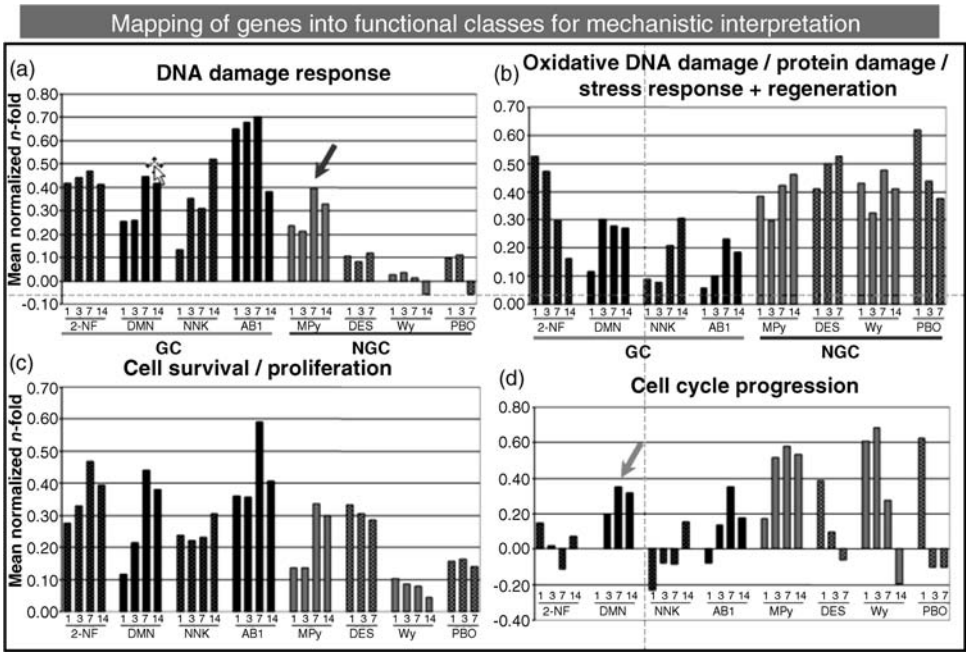


Figure 6.1 Major biological processes affected by the two major classes of carcinogens, as analyzed for the GCs 2-NF, DMN, NNK, and AB1, and the NGCs MPy, DES, Wy, and PBO. To calculate the normalized *n*-folds for these diagrams, the expression of each single gene (average fold deregulation versus time-matched control average) was normalized over all samples by dividing the value in a particular sample through the maximum deregulation found for that gene in all samples, resulting in fold deregulation values from −1 to 1. Then the mean of the normalized values was calculated for each single sample from all genes belonging to the toxicological categories indicated above each diagram. Thus, this mean represents the overall response of the genes representing a particular pathway or pathway combination. (Modified from Figure 3 in Ref. [24].)

observed *in vitro*, when cells containing wild-type p53 were treated with a comparable selection of GCs for just 4 h [26,27]. A "survival" response as mentioned above for NNM was seen with all four liver GCs, yet clearly with different strength of expression deregulation (Figure 6.1). The four liver GCs showed additional gene expression changes, which were mostly compound specific, for example, increased expression of genes indicating inflammation following the necrosis induced by DMN (see Figure 1 in Ref. [23]).

In another study, the liver expression profiles induced by four liver GCs (2-acetylaminofluorene (2-AAF), 2,4-diaminotoluene (2,4-DAT), 2-nitropropane (2-NP), and 2-nitro-*p*-phenylenediamine (2-NpP)) after 28 days of treatment were measured on a custom oligonucleotide microarray and compared with those induced by structurally similar, yet noncarcinogenic isomers (4-acetylaminofluorene (4-AAF), 2,6-diaminotoluene (2,6-DAT), 1-nitropropane (1-NP), and 4-nitro-*o*-phenylenediamine (4-NoP)) [28]. Cluster analysis with significantly

deregulated genes, previously selected with statistical and ratio cutoffs between treated and time-matched control samples, revealed the following pattern: a cluster of the three clear GCs 2-AAF, 2,4-DAT, and 2-NP, and a looser cluster of all four isomers. 2-NpP grouped with the noncarcinogenic isomers, which was not completely unexpected since this compound appears to induce liver cancer only in female mice, but not in rats of both sexes. Several of the genes upregulated specifically by the GCs were again Mgmt and p53 target genes such as p21and Ccng1, in agreement with the studies mentioned above for GCs, and many other studies not cited here. Overall investigation of gene expression changes in rodent liver after treatment with GCs defined a clear DNA damage response with rather short-term induction of several characteristic genes and suggested in addition a positive effect on "survival" signaling by rodent genotoxic hepatocarcinogens.

Quite a few studies evaluated NGC-induced profiles in rat and mouse livers, sometimes in comparison with GCs and/or noncarcinogens. In one of the early studies [29], B6C3F1 mice were treated with oxazepam and Wyeth-14,643 (Wy), which are both nongenotoxic rodent hepatocarcinogens. Iida *et al.* [29] summarize evidence that both induce oxidative stress that may be involved in tumor induction. Gene expression analysis after 2 weeks and 6 months of treatment with a cDNA array revealed induction of, for example, mRNAs encoding phase I biotransformation enzymes of the cytochrome P450 2b and 4a family and the stress-responsive growth arrest and DNA-damage-inducible, beta (GADD45b) mRNA by both compounds. Yet there were also several mRNAs increased by oxazepam or Wy only, letting the authors to suggest that the early carcinogenic pathways between these compounds are different [29]. Extending this work, gene expression was analyzed in mice that were treated for 2 weeks with three known carcinogens, including oxazepam, *o*-nitrotoluene, and methyl eugenol, and the noncarcinogens *p*-nitrotoluene, eugenol, and acetaminophen [30]. Again, only few genes were commonly deregulated by the carcinogens, although there was some resemblance between *o*-nitrotoluene and methyl eugenol, which both induced p21 and Ccng1, whereas oxazepam was clearly different at least at the early time point examined. Concerning the noncarcinogens, eugenol induced some of the same genes as methyl eugenol, although with a lower deregulation ratio, but did not lead to downregulation of two genes strongly decreased by methyl eugenol, which may indicate reduced apoptosis in case of methyl eugenol. Decreased apoptosis is in fact a mechanism supposed to be involved in cancer induction. The reason for the quantitative differences in the effects of methyl eugenol and eugenol may be a more efficient metabolism in case of the latter, thus suggesting the existence of a threshold for cancer induction in this compound class. Differences in metabolism were also suggested for the pair *o*-nitrotoluene and *p*-nitrotoluene to explain why the latter does not induce liver tumors [30]. Although in these studies the genes deregulated by the carcinogens only partially overlapped, there were indications that cell cycle and apoptosis functions were affected by all carcinogens already 2 weeks after treatment. Overall, these first studies with different carcinogens also suggest that clearly defined

carcinogens at least with respect to the major mechanisms, that is, GC versus NGC, are needed to be able to select gene groups suggesting functions that may be disturbed early on in the progress of cancer induction.

To investigate whether rodent liver NGCs, similar to GCs, deregulate certain sets of genes or at least affect similar functions, which can be deduced from the affected genes, gene expression patterns were investigated in liver after treatment of rats with several rodent liver NGCs, including methapyrilene hydrochloride (MPy), Wy, diethylstilbestrol (DES), and piperonyl butoxide (PBO), in comparison with the GCs 2-NF, DMN, NNK, and AB1 [24]. Characteristics of the gene deregulations by four of these GCs were described earlier. NGC-induced expression profiles represented cell cycle progression, pathways likely involved in regeneration such as protein synthesis, and responses to oxidative DNA or protein damage. These functional alterations, indicated by the gene expression changes, showed different time course patterns depending on the NGC class. For example, upregulation of cell cycle progression genes either was transiently induced or increased with time, likely reflecting either direct mitogenic effects or regenerative hyperplasia following cytotoxic cell damage. The former was observed for Wy, DES, PBO, and also TAA, and the latter for MPy (Figure 6.1). The MPy-associated gene expression profile also suggested induction of apoptosis and inflammation, as expected for a compound inducing regenerative hyperplasia. The different NGCs were overall associated with different combinations of affected functions and pathways, in agreement with the various mechanisms suggested to contribute to cancer induced by NGCs. In addition, the NGCs also showed compound-specific gene expression changes that may or may not contribute to carcinogenicity. Gene expression analysis thus allowed to molecularly characterize early events following NGC exposure, including at least some of the suggested mechanisms. They also revealed that MPy as an NGC did induce a partial DNA damage response (Figure 6.1a) that may be explained by oxidative or other stress secondarily leading to DNA damage. On the other hand, DMN as a GC also increased cell cycle progression (Figure 6.1d) starting at day 7 after treatment. This can be explained by regenerative hyperplasia since DMN also induced necrotic damage to the liver.

A time-dependent increase in cell cycle progression in the liver after treatment of rats with NGCs was confirmed by searching the Ingenuity Pathway Analysis (IPA) knowledgebase for over- and underrepresented biofunctions within the gene expression data. Furthermore, activated or inhibited transcription factors predicted from the gene expression profiles of their known target genes suggested activation of the tumor suppressor gene TP53 by GCs, corresponding to a DNA damage response directly activated by this class of carcinogens. For NGCs, the transcription factors Myc and FoxM1 were predicted as activated. Both transcription factors are known to induce transcription of various genes involved in cell cycle progression [31–33].

Overall, the characteristic profiles obtained after short-term treatment could discriminate genotoxic and nongenotoxic rodent liver carcinogens from noncarcinogens, at least for the compounds investigated.

A preliminary comparison of the gene expression changes induced in the rat liver by carcinogens with those induced in the kidney (unpublished) by two kidney GCs and one kidney NGC revealed that the kidney GC signature at the gene level is rather similar to the liver GC signature concerning the DNA damage response. The kidney NGC also showed induction of cell proliferation at the gene expression level with the time dependence of deregulation reflecting the primary mechanism. Although other responses were not as conserved at the gene level, they still were reflected in the gene expression changes, for example, regeneration and dedifferentiation, which one would expect to employ different genes specific for the differentiation state of a certain organ. Similar to these findings with liver and kidney gene expression profiles, Thomas *et al.* also observed that certain gene expression profiles associated with carcinogen exposure are conserved across several tissues [13].

6.2.2 Approaches for Prediction of Potential Hepatocarcinogens Based on Gene Expression Profiling

Already early on, when transcriptomics was applied in toxicological studies, it was suggested that gene expression profiles measured after short treatment with compounds may predict their chronic toxicity effects, including, for example, a potential for induction of carcinogenicity. This was based on the general assumption that toxicants induce characteristic early molecular changes that are reflected in their gene expression responses. It was also based on data of early toxicogenomic investigations in the hepatotoxicity area, which, in addition to the studies described above, had shown that, for example, rat liver expression profile clustering could distinguish between hepatotoxins with different mechanisms of toxicity after 3 days of treatment [34], and 1-, 3-, and 14-day rat liver mRNA profiles could classify samples derived from rats exposed to enzyme inducers versus those exposed to peroxisome proliferators [35]. Based on up to 14-day liver expression profiles after treatment with 25 different compounds and matched controls, Steiner *et al.* [36] could discriminate between hepatotoxic and nonhepatotoxic compounds with 90% true positive prediction, and even between four classes of liver toxins (cholestasis, steatosis, and direct-acting peroxisome proliferator-activated receptor α (PPAR-α) agonists) with reasonable accuracy.

Several limitations were discussed by Steiner *et al.* [36], which in principle apply to all toxicity predictions based on compound-induced gene expression, especially when complex toxicities are considered. For calculation of the prediction accuracies, so-called external cross-validation was applied by Steiner *et al.*, which removes expression profiles representing all samples derived from one compound at a time, uses the remaining as a training set to calculate a classifier, and then predicts the left out samples, that is, one compound. After iterative cycles, this then estimates the prediction accuracy of the classifier. Although this approach is still less biased with respect to performance estimates than, for example, internal cross-validation, it is still too optimistic and not as stringent as

prediction of an independent test set. Also, to train classifiers for complex toxicities such as hepatotoxicity, which applies even more to nongenotoxic carcinogenicity, the compound database should cover all possible mechanisms. Nevertheless, since prediction of especially nongenotoxic carcinogenicity with short-term assays would offer new opportunities concerning carcinogenicity testing, and since theoretically such a prediction should be conceivable if the profiles cover important mechanisms and relevant doses and time points, prediction approaches for carcinogens were evaluated by several groups (Table 6.1). Major efforts will be summarized in the following.

In one of the first studies, 1-day expression profiles of rats treated with 24 NGCs and 28 noncarcinogens were profiled on cDNA arrays representing 1471 genes selected by literature searching and prior performance [37]. The 24 nongenotoxic carcinogens included some targeting organs other than the liver. Doses used were in general highly toxic ones corresponding to 30–50% LD_{50} (lethal dose 50%) values. Genes were selected by a univariate *t*-test between treated and control groups, and by multivariate methods employing several algorithms. The genes from the different analyses were combined and subjected to a 10% cross-validation, meaning that at each iteration 10% of the samples were used as a test set. This yielded six genes that identified NGCs with 88.5% prediction accuracy estimated by cross-validation, including "correct" prediction of several extrahepatic NGCs. The available gene expression profiles were also analyzed for the molecular mechanism of nongenotoxic carcinogenicity using the 125 genes derived from the univariate *t*-test. This revealed connections of the deregulated genes to the c-Myc proto-oncogene and to fibronectin (FN)-regulated, integrin signaling pathways [37]. A limitation to consider here is the rather high dose used in comparison with the 2-year cancer study, which may overestimate the NGC potential for a compound; that is, at this dose animals would not survive for 2 years, whereas at lower doses, tumors may not be induced. The prediction of extrahepatic NGCs with liver gene expression data was explained by the fact that in some cases disturbed metabolism in the liver may increase or decrease levels of metabolites or hormones that then exert nongenotoxic tumor-inducing effects in other organs. Although this is true in some cases, one may also interpret this result as nonspecific.

In a similar rather short-term approach, Fielden *et al.* [38] evaluated rat liver expression profiles after an up to 7-day exposure to rat liver NGCs and nonhepatocarcinogens using the "CodeLink" RU1 BioArray containing 30-mer probes for 9911 rat genes. Compounds were classified as rat liver NGCs or noncarcinogens based on clear rules, with NGCs, for example, being positive in the 2-year bioassay in at least one rat strain or gender and believed to induce cancer via a nongenotoxic mechanism according to available data, and with noncarcinogens being negative in the 2-year bioassay in either gender of rats [38]. Profiles for 147 compounds were available in the toxicogenomic reference database DrugMatrix [44], of which 25 NGCs and 75 nonhepatocarcinogens were used as a training set, and the remaining 47 chemicals (21 positives and 26 negatives) set aside as a test set. A discriminating gene signature was derived from 3-, 5-, or 7-day

Table 6.1 Summary of toxicogenomic approaches for prediction of a potential for tumor induction based on gene expression profiling.

Publication/ signature	Species, strain, sex, group size	Training set compound classes	Test set compound classes	Tumor target organ	Treatment durations	Doses	Technology platform	Signature gene sets	Prediction accuracy	Comments (per publication)
Nie *et al.* [37]: J&J data set	Rat, SD, male, 3	24 NGCs (some inducing tumors in organs other than liver: extrahepatic NGCs)	No extra test set included	Liver mostly	1 day	30–50% LD_{50}	cDNA array	J&J – 6 genes	Training set cross-validation: 88.5% accuracy	Includes correct prediction of extrahepatic NGCs, may be explained by altered metabolism of, for example, hormones in the liver, which affect other organs
		28 non-carcinogens								*Limitations*: high dose used in comparison with 2-year bioassay may overestimate an NGC potential
Fielden *et al.* [38]: Iconix data set	Rat, SD, male, 3	25 NGCs (positive in the 2-year bioassay in at least one rat strain or gender)	21 NGCs	Liver	Training set: 3, 5, and 7 days Test set: 1, 3, and 5 days	MTD selected in initial range finding studies: 5–10% reduction of BW gain	Amersham CodeLink array RUI	Iconix – 37 genes	Training set cross-validation: 56% sensitivity, 94% specificity	Functional annotation of the 37 signature genes: pro-proliferative and anti-apoptotic effects
		75 nonhepatocarcinogens (in the 2-year bioassay in either gender of rat)	26 non-hepatocarcinogens						Test set prediction: 86% sensitivity, 81% specificity	Clustering of samples based on the 37 signature genes: four clusters, suggested by the authors to represent at least four distinct modes of action of liver tumor formation, that is, toxicity and regeneration effects, hormonal effects, xenobiotic receptor-mediated effects, and peroxisome proliferators
										Limitations: the 37-gene profile of the cluster groups does not easily allow to deduce corresponding modes of action

Fielden *et al.* [39]: combination of several data sets	Rat, SD, male, 3	J&J data set	Iconix: 30 NGCs, 31 negative GSK: 25 NGCs, 53 negative	Liver	Training set: 1 day Iconix test set: 3, 5, and 7 days GSK test set: 4 days	Training set: 30–50% LD_{50} Iconix test set: MTD at 5 days GSK test set: MTD at 4–5 days	J&J training: cDNA array Iconix test: Affx RG230_2.0 GSK test: Affx RAE230A	J&J – 6 genes	*Test set prediction*: Iconix: 53% sensitivity, 74% specificity GSK: 96% sensitivity, 36% specificity	Samples from compounds overlapping with the training set were excluded
	Rat, SD, male, 3	Iconix data set	J&J: 8 NGCs, 24 negative GSK: 25 NGCs, 53 negative	Liver	Training set: 3, 5, and 7 days J&J test set: 1 day GSK test set: 4 days	Training set: MTD at 5 days J&J test set: 30–50% LD_{50} GSK test set: MTD at 4–5 days	Iconix training: CodeLink RU1 J&J test: CodeLink whole genome GSK test: Affx RAE230A	Iconix – 37 genes	*Test set prediction*: J&J: 100% sensitivity, 63% specificity GSK: 73% sensitivity, 60% specificity	Sensitivity and specificity: dependent on both the signature and the test set, appears somewhat more stable with the Iconix signature derived from up to 7-day treatment and containing more genes
Fielden *et al.* [40]: PSTC effort to reevaluate former RNA samples with qPCR	Training set: rat, SD, male Test set: rat, SD or F433, mostly male, ≥3	Iconix: 23 NGCs, 49 non-hepato-carcinogens RNA samples derived from several laboratories	15 NGCs 51 non-hepato-carcinogens RNA samples derived from several laboratories	Liver	Iconix training set: 1, 3, 5, and 7 days	Iconix training set: MTD at 5 days Test set: equal to or higher than those used in the 2-year bioassay	TaqMan qPCR on array cards	Final qPCR set – 22 genes Derived from: Iconix – 11 genes J&J – 6 genes Ellinger – 10 GC genes	Overall test set prediction: 67% sensitivity, 59% specificity	Test samples are from a variety of studies performed at several laboratories, several of which were derived from different study designs, rat strains, and/or compound classes in general Reduced sensitivity/specificity were driven by the J&J and the NTP data sets. Samples for both data sets were derived from different overall study designs. The NTP data [41] were from F344 rats, and employed primarily nontherapeutic chemicals that may work via modes of action not reflected in the training data set

(*continued*)

Table 6.1 (*Continued*)

Publication/ signature	Species, strain, sex, group size	Training set compound classes	Test set compound classes	Tumor target organ	Treatment durations	Doses	Technology platform	Signature gene sets	Prediction accuracy	Comments (per publication)
Ellinger *et al.* [22]	Rat, Wistar, male, 3	5 GCs (2 TPs per compound) 5 NGCs (2 TPs per compound) 3 non-carcinogens (4 TPs per compound)	4 GCs 6 NGCs 6 non-carcinogens Several time points each	Liver	1, 3, 7, and 14 days	Based on those reported to induce liver tumors in the 2-year rat bioassay	Affx RAE230A	#1: mechanistic – 141 genes #2: RFE – 256 genes #3: SVM – 512 genes #4: ANOVA – 2048 genes	#1: 81% accuracy (test set) #2: 88% accuracy (test set) #3: 88% accuracy (test set) #4: 75% accuracy (test set)	Exact time points for training were selected from a prior mechanistic analysis of the gene expression profile induced by these compounds [24] GCs were included to distinguish their mode of action represented in the gene expression profiles [23] from those suggested for the NGCs Mechanistic evaluation of the discriminant genes (all groups), overrepresentation of • Oxidative stress/DNA damage response, characteristic for GCs • Cell proliferation, characteristic for NGCs
Uehara *et al.* [42]: TG-GATEs data set	Rat, SD, male, 5	6 NGCs, associated with hepatocellular necrosis/degeneration (HD, 28 days) 54 non-hepatocarcinogens (LD, MD, 28 days)	(1) Necrotic NGCs: low and medium doses and TPs <28 days (2) GCs (3) NGCs: associated with liver enzyme induction, peroxisome	Liver	3, 6, and 9 h, 1, 3, 7, 14, and 28 days	Low (L), medium (M), and high (H) doses High dose: MTD at 7 days	Affx RG230_2.0	Final set – 9 genes	Training set cross-validation: 99% sensitivity, 97% specificity	Test set samples predicted as positive (i.e., as "necrotic NGC"): (1) Necrotic NGCs: HD at earlier TPs, MD at 28 days (2) 3 rodent liver GCs, TPs ≥15 days None of the NGCs in the test set with a non-necrotic mode of action were predicted as positive

			proliferation, hormonal perturbations (4) Non-hepatocarcinogens								Limited mechanistic analysis with 82 genes derived from the first analysis step: networks around the stress-responsive kinase p38 and the proto-oncogene c-Myc
Auerbach *et al.* [41]: environmental chemicals including alkylbenzenes	Rat, F433, male	4 carcinogens (3 Ames-positive, i.e., GCs) 3 nonhepatocarcinogens	Alkylbenzenes (all Ames-negative): 3 carcinogens 3 nonhepatocarcinogens 2 not yet tested for carcinogenic properties	Liver	2, 14, and 90 days	Training set: 1 dose level Test set: 2 dose levels	Agilent rat whole genome 44k	20 genes, selected as informative by several SVM models	Training set cross-validation: mostly 100% accuracy for SVM models from combinations of profiles derived from different TPs Test set prediction: 53–90% accuracy, depending on the exact model and exposure duration	The authors suggest that exposure times ≥90 days are required to induce gene expression changes in the liver that are clearly indicative of hepatocarcinogenic potency
											Mechanistic evaluation in general • Cell proliferation, DNA repair, apoptosis increased by carcinogens after 14 days • Xenobiotic metabolism increased at 90 days

(*continued*)

Table 6.1 (*Continued*)

Publication/ signature	Species, strain, sex, group size	Training set compound classes	Test set compound classes	Tumor target organ	Treatment durations	Doses	Technology platform	Signature gene sets	Prediction accuracy	Comments (per publication)
Thomas *et al.* [43]: lung carcinogens, environmental/industrial chemicals	Mice, B6C3F1, female, 3–5	13 carcinogens (GCs, NGCs) 12 non-carcinogens (lung) 1 equivocal	No extra test set included	Lung	13 weeks	Carcinogenic dose	Affx Mouse Genome 430 2.0		Training set cross-validation: 77.5% average accuracy, generally higher specificity than sensitivity	Endpoint "increased lung tumor incidence" in the mouse 2-year bioassay. Chemicals: rather diverse concerning route of exposure, chemical structures, results from genotoxicity tests, that is, represented both GCs and NGCs that were classified all as carcinogens Broad selection of feature selection and classification algorithms tested in conjunction with a range of cross-validation conditions

Abbreviations: GC, genotoxic carcinogen; NGC, nongenotoxic carcinogen; SD, Sprague Dawley; BW, body weight; LD_{50}, lethal dose 50%; PSTC, Predictive Safety Testing Consortium; Affx, Affymetrix; TP, time point; LD, low dose; MD, medium dose; HD, high dose; RFE, recursive feature elimination; SVM, support vector machine; ANOVA, analysis of variance.

profiles with a certain linear model algorithm (adjusted sparse linear programming (A-SPLP)), which simultaneously trained a classifier. Functional annotation of the 37 signature genes including consideration of their direction of deregulation indicated pro-proliferative and anti-apoptotic effects as expected for NGCs. Cross-validation within the training set estimated a sensitivity of 56% and a specificity of 94%. Application of this classifier on the test set profiles yielded a sensitivity of 86% and a specificity of 81%. This was compared with other parameters; for example, increased liver weight and serum ALT had a similar (81%) or even higher (89%) specificity, respectively, but a lower sensitivity of 48 and 57%. Based on clustering of the expression of the 37 signature genes into four clusters, the authors suggested that at least four distinct modes of action of liver tumor formation were represented, that is, toxicity and regeneration effects, hormonal effects, xenobiotic receptor-mediated effects, and peroxisome proliferators. Although the compounds clearly clustered in this way, it is noted that the 37-gene profile of these cluster groups as visualized with a heat map did not appear to be clearly different between these compound groups to be able to deduce corresponding modes of action. The clustering was based on impact scores, which also considered a certain weight added to each gene. One would need to apply other gene selection and classifier algorithms to check the stability and thus importance of the signature genes and classifier in this data set.

In an effort to evaluate a genomic signature for prediction and mechanistic assessment of rat liver NGCs, the Predictive Safety Testing Consortium (PSTC) of the Critical Path Institute made use of the two data sets described above abbreviated as the J&J [37] and Iconix [38] data sets and signatures [39]. Each data set was classified with the other signature, and in addition a data set from GSK was used as a test set for both signatures. The latter was derived from a 1-day and a 4-day treatment of rats. Compounds were excluded from test sets when they overlapped with compounds used for the respective training sets. Adjustments of the expression data were also made to account for differences in the dynamic range of the different microarray platforms used. Thus, the 37-gene Iconix signature was used for the classification of the J&J and GSK test sets, and the 6-gene J&J signature was employed for the classification of the Iconix and GSK test sets. The accuracies of the predictions, and especially sensitivity and specificity, were clearly dependent on both the signature and the test set. The 37-gene Iconix signature resulted in relatively higher sensitivities for both the J&J and GSK test sets (100 and 73%, respectively) compared with specificity values (63 and 60%, respectively), whereas the 6-gene J&J signature showed relatively higher specificity of 74% (versus 53% sensitivity) with the Iconix data set, but a completely different picture with only 36% specificity and 96% sensitivity with the GSK data set. These data, being derived from different laboratories using different study designs, microarray platforms, and processing protocols, represent realistic performance values of the two signatures.

Although the overall accuracy of 72% for one combination (Iconix signature, J&J test set) indicated significant classification accuracy, prediction accuracies for the other combinations were lower (down to 55%). Reasons for these

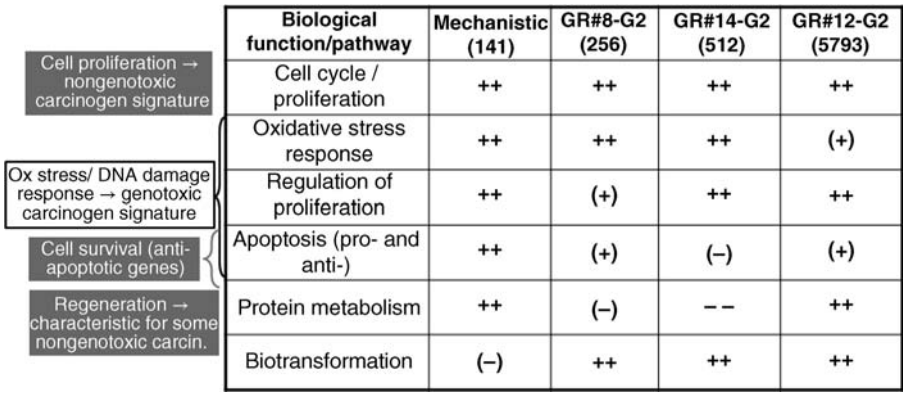
discrepancies are likely differences in the animal study design, microarray platform, and protocol, data processing and analysis, and classification of the compounds into positive and negative samples with respect to, for example, some positive findings. To control at least for the gene expression measurement platform and protocol, it was decided to rederive a signature on a qPCR platform, which is in general more cost effective and standardized and may be more readily accessible than a microarray platform. The objective of this further effort was not to replace the 2-year bioassay, but rather to guide internal decision making, allow prioritization of chemicals for chronic testing, and/or enable a more rapid understanding of the mode of action [40]. The selected gene set for the qPCR included 11 genes from the 37-gene Iconix signature, the 6-gene J&J signature, 3 normalizer genes, and 10 genes from the GC signature described previously [23] to distinguish genotoxic from nongenotoxic modes of action. To develop the qPCR signature, RNAs from the Iconix rat liver samples representing 72 compounds after 1-, 3-, 5-, and 7-day treatments were reanalyzed. For development of the qPCR signature and the classifier model, samples from all 72 compounds (23 NGCs and 49 nonhepatocarcinogens) were used after an initial evaluation of a model developed from training and test sets of these 72 compounds. This model was then tested on an independent test set representing 15 rat liver NGCs and 51 nonhepatocarcinogens. The test set contained a broad array of samples from a variety of studies performed at several laboratories, several of which were derived from different study designs, rat strains, and/or compound classes in general. The qPCR signature developed with the 72 compounds retained 22 of the original 27 genes, with peptidylprolyl isomerase A as a normalizer gene. Applying the classifier model with the 22-gene signature on the test set resulted in 67% sensitivity and 59% specificity [40]. Such sensitivity may be acceptable, depending on the use of the prediction, yet the specificity was rather low. Detailed examination of the contribution of the different data sets to these values indicated that the reduced sensitivity and reduced specificity were driven by the J&J and the NTP data sets. Samples for both data sets were derived from rather different overall study designs. For example, the 1-day only and rather high-dose exposure of the J&J data [37] may have introduced variable false-positive gene expression changes that after a few more days may have been compensated by protective responses. The NTP data [41] were derived from F344 rats, in contrast to Sprague Dawley rats used in all other cases, and employed primarily nontherapeutic chemicals that may work via modes of action not reflected in the training data set derived mostly from therapeutic compounds. Concerning the performance of the qPCR and the microarray models, no major differences were observed.

Overall, this evaluation confirmed the suggestions that for a classification approach test samples should be generated with protocols rather similar to those used for the training set and that although short-term treatments can be used, they should include periods of 3 days or more, when the very early gene expression changes have been stabilized or become compensated. It is also recommended to check the signature for mechanistic content. Although the signature

genes are selected for optimal prediction performance, they should also represent the underlying biology of the compound classes. The final 22 qPCR signature developed here, despite not representing complete pathways, did contain genes that by literature mining and pathway analyses are associated with pathways suggested to be involved in carcinogenesis, such as growth regulation, DNA repair, and oxidative stress [40].

The authors mention an important point, which should be considered in any prediction exercise: The quality of prediction models for a certain phenotype depends not only on the applicability of the technology and study design of the training samples, but also on the accurate classification of the samples with respect to the phenotype to be modeled. Concerning rodent hepatocarcinogenicity, the outcome of the 2-year bioassay also shows some variability depending on several study factors, such as strain and gender. Thus, the accuracy obtained for the genomic signature is determined not only by the intrinsic variation in the gene expression assay, but also by the variation in the benchmark phenotype that is to be modeled by the gene expression assay. Therefore, the sensitivity and specificity values calculated are not an absolute measure of performance, but rather should be seen as a composite estimate for these measures. In this respect, a predicted potential to induce cancer via a NGC mode of action based on compound-induced gene expression changes always needs to be evaluated in the context of dose response, pathological changes, results from genotoxicity testing, and any mechanistic data.

A further attempt to derive gene expression signatures for classification of compounds with respect to their potential to induce cancer in the rat liver also incorporated mechanistic considerations as a basis to evaluate the signatures [22]. With a training set of five GCs (two time points per compound), five NGCs (two time points per compound), and three noncarcinogens (four time points per compound), a three-class classifier was derived from rat liver gene expression profiles after up to 14 days of treatment with three replicate animals per exposure group. The rationale for dose selection was based on those reported to induce liver tumors in the 2-year rat bioassay, and the exact time points (two per carcinogen) were selected from a prior mechanistic analysis of the gene expression patterns induced by these compounds. GCs were included to distinguish their mode of action clearly represented in the gene expression profiles [23] from those suggested for the NGCs, and the exact NGCs included were compounds likely associated with different NGC modes of action [24]. Three different algorithms, including recursive feature elimination (RFE), support vector machine (SVM), and analysis of variance (ANOVA), were used to determine three sets of discriminant genes, abbreviated as #2(256)RFE, #3(512)SVM, #4(2048)ANOVA, which all yielded similar cross-validation accuracies [22]. These three and an additional mechanistic set of 141 genes representing pathways and functions previously found to distinguish between several rodent liver GCs and NGCs [24] were then employed to classify an independent test set of four GCs, six NGCs, and six noncarcinogens, with samples derived from four time points per compound (1, 3, 7, and 14 days). With compounds assigned to



Biological function/pathway	Mechanistic (141)	GR#8-G2 (256)	GR#14-G2 (512)	GR#12-G2 (5793)
Cell cycle / proliferation	++	++	++	++
Oxidative stress response	++	++	++	(+)
Regulation of proliferation	++	(+)	++	++
Apoptosis (pro- and anti-)	++	(+)	(–)	(+)
Protein metabolism	++	(–)	– –	++
Biotransformation	(–)	++	++	++

Figure 6.2 Biochemical functions/pathways represented by the gene groups used for prediction. Overrepresented functions/pathways in the gene groups used for predictions [22] as determined by a Fisher's exact test. (IPA software) Over- or underrepresentation is indicated by + and −, respectively, and the degree thereof by parentheses, or one or two signs. Effects at the cell and/or tissue level, which result from alterations of these pathways and which are characteristic of one or the other carcinogen class, are indicated on the left.

either the GC or NGC class even when just one sample per compound was classified as such, prediction accuracies ranged from 75 to 88%, depending on the gene set used (Table 6.1). For a mechanistic evaluation of the discriminant genes, a Fisher's exact test was applied to extract over- and underrepresented biological pathways/functions associated with the genes in these groups (Figure 6.2). At least an oxidative stress/DNA damage response, which is characteristic for GCs, and cell proliferation, which is characteristic for NGCs, were encoded by all gene groups irrespective of how exactly they were selected. Genes encoding phase I, II, and III biotransformation enzymes and proteins were also found overrepresented in the gene lists derived from gene ranking. They may distinguish these two classes of carcinogens based on the fact that the GCs appear to upregulate a defined set of such genes, whereas the NGCs affect the corresponding mRNAs in a more compound-specific manner. Thus, the discriminant genes do represent biological processes that according to current knowledge likely contribute to early cancerogenesis.

Interestingly, ethionine, which is mentioned in the literature as NGC, since it is negative in the Ames test, was predicted as GC based on expression profiles of rat liver samples after 3 days of treatment, likely due to induction of an observed DNA damage response at the gene expression level. Based on current understanding, the ethionine prediction obtained here had to be interpreted as false positive. Although being negative in the Ames test, ethionine was reported to be positive in the *in vivo* micronucleus assay, a genotoxicity assay detecting genome and chromosomal mutations. Such a positive result should by no means be taken as strong evidence for direct genotoxicity, yet together with the induced gene expression pattern it may indicate a genotoxic mode of action in the rat liver, potentially due to a metabolite generated in this organ. In fact, *S*-vinyl

homocysteine was suggested as a possible proximate carcinogenic metabolite already in 1979 [45], since *S*-vinyl homocysteine is highly mutagenic for the TA100 strain of *Salmonella typhimurium*. Overall, the results derived from the gene expression data including the hint from classification and especially a closer investigation of the gene expression changes in a mechanistic sense suggest that the carcinogenic mechanism of ethionine needs further clarification.

The Toxicogenomics Project in Japan developed a Genomics-Assisted Toxicity Evaluation System (TG-GATEs), which is based on gene expression profiles derived from livers of rats after treatment with hepatotoxicants at three different dose levels for 3, 6, and 9 h, and for 1, 3, 7, 14, and 28 days [42] (see also Section 6.2.4). The compound list contained several carcinogens of different classes. To derive a model for rodent liver NGCs, profiles of six NGCs known to be also associated with hepatocellular necrosis/degeneration (high dose, 28 days) were used as positive training "necrotic NGC" data, with 54 nonhepatocarcinogens (low and middle doses, 28 days) as a negative training set. The test set included (1) the remaining middle and low doses and earlier time points of the positive necrotic NGCs, (2) GCs, (3) NGCs whose carcinogenic mode of action is suggested to be related to liver enzyme induction, peroxisome proliferation, and hormonal perturbations, and (4) nonhepatocarcinogens. A classifier training approach using SVM combined with a certain gene selection method yielded 82 top-ranked genes for differentiation between necrotic NGCs and nonhepatocarcinogens, which were reduced to a signature of nine genes using fivefold cross-validation [42]. The derived nine-gene classifier was associated with a sensitivity and specificity of 99 and 97%, respectively. It is to be kept in mind that these are the data from cross-validation within the training set, which represents 6 positive (high dose!) and 54 negative (middle and low dose!) compounds, and do not represent an independent test set. Test set samples predicted as positive were several of the other doses and time points of the "necrotic" NGCs, including middle-dose samples after 28 days of treatment with two of these compounds, and high-dose samples after 8 and 15 days (two NGCs) or 15 days (one NGC) of treatment. None of the low-dose or 1- and 3-day samples of the necrotic NGCs were predicted as positive with the nine-gene classifier. Furthermore, none of the other NGCs in the test set with a non-necrotic mode of action were predicted as positive. However, test set samples from the longer treatments of three rodent liver GCs were classified as "necrotic" carcinogens. The authors suggest that although the primary mode of action between GCs and NGCs is different, gene expression profiles after 28 days of treatment with GCs and necrotic NGCs may reflect commonly affected processes. A preliminary pathway analysis with the 82 discriminant genes derived from the first step selected networks around the stress-responsive kinase p38 and the proto-oncogene c-Myc, with an extensively described role in tumorigenesis at least for the latter.

The TG-GATEs approach with a training NGC group of a rather closely defined mode of action and the predictions obtained for different carcinogen classes highlights that one must be very clear what is being modeled, including a close consideration of the chosen time points and gene deregulations with their

represented functions observed at these time points, to be able to interpret the classifier output. As pointed out earlier, it is therefore necessary to perform a functional interpretation of the signature genes. The nine genes of the final signature appear too few to derive a clear mode of action hypothesis, so this exercise was performed with a prior gene list selected in the first step. Overall, the data may be enhanced by a more detailed prior mechanistic analysis of the gene expression profiles over time, which are associated with the different classes of NGCs in comparison with the GCs.

A huge number of chemicals are present in the environment that should be assessed concerning their carcinogenic properties. Up to now, the National Toxicology Program (NTP) has been able to test about 1485 chemicals using the 2-year bioassay with a standard NTP protocol based on F344 rats. Yet many more chemicals would need to be tested; for example, over 75 000 chemicals are listed in the US EPA's Toxic Substances Control Act Inventory (summarized in Ref. [41]). Considering these numbers, it is obvious that efficient methods are needed to determine the carcinogenic potential of these many chemicals, including methods to prioritize them for chronic testing. Therefore, SVM classification models build with liver gene expression profiles of male F433 rats after treatment with hepatocarcinogens and nonhepatocarcinogens for 2, 14, or 90 days were evaluated on an independent test set of alkylbenzene flavoring agents with a range of hepatocarcinogenic properties. The training set included different chemical structures and represented three nonhepatocarcinogens and four carcinogens, of which three were Ames-positive GCs. The test set, however, included only alkylbenzenes, three of which were shown to induce and three not to induce rodent liver cancer, and two of which were not yet tested for their carcinogenic properties. All of the alkylbenzenes in the test set were Ames-negative. Methyl eugenol, one of the alkylbenzenes, was used at a higher dose as a training sample and at a lower dose as a test sample. The different SVM models were built using combinations of profiles derived from different time points and resulted in models with almost only 0% cross-validation errors. For the independent test set, error rates ranged from 10 to 47%, depending on the exact model and the exposure duration of the test compound. With the models calculated for this data set, performance increased with increasing exposure duration. An analysis of the pathways that were affected at the different time points suggested cell proliferation, DNA repair, and apoptosis to be increased after 14 days of exposure to carcinogens, whereas at the 90-day time point genes related to xenobiotic metabolism showed higher expression in the carcinogen-derived samples. From their results, the authors suggest that longer exposure times, for example, 90 days or more, are required to induce gene expression changes in the liver that are clearly indicative of hepatocarcinogenic potency [41]. A possible interpretation of this apparent contrast to other classification approaches summarized here, which in several cases delivered reasonable predictions accuracies with shorter treatment durations, will be given the end of this chapter.

Quite a few environmental and/or industrial chemicals induce lung cancer in mice, as observed in NTP studies using B6C3F1 mice. As mentioned earlier,

since huge numbers of chemicals are still untested for their cancer-inducing characteristics, more efficient ways of safety testing are required, as, for example, listed in the report of the US National Research Council (NRC) on "Toxicity Testing in the 21st Century" [46]. To evaluate whether a shorter term alternative to the standard 2-year bioassay would be feasible, researchers in the environmental toxicity area investigated whether gene expression changes in mouse lung obtained after a 13-week treatment could predict increased lung tumor incidence [43]. A 13-week and not shorter treatment duration was chosen based on the rationale that earlier the gene expression changes induced by different carcinogen classes, especially NGCs, would be rather transient and thus less robust, whereas over, for example, a 13-week time course, gene expression changes would evolve to a profile comparable between carcinogens irrespective of primary class. A total of 26 chemicals were profiled over a period of 3 years, of which approximately half had produced lung tumors in mice. The chemicals were rather diverse concerning route of exposure, chemical structures, and results from genotoxicity tests, that is, represented both GCs and NGCs that were all classified as carcinogens. Batch effects, which are inevitable in microarray studies when testing is performed over longer periods of time [47], were removed based on matched controls run in parallel to treatments. For calculation of classification models with the endpoint "increased lung tumor incidence" in the mouse 2-year bioassay, a broad array of feature selection and classification algorithms was tested in conjunction with a range of cross-validation conditions. The average accuracy obtained for the best-performing models was 77.5% with generally higher specificity than sensitivity, that is, less prediction of false positives than false negatives. This study also investigated how many compounds are likely required to produce a robust classification model from their associated gene expression profiles. A learning curve analysis revealed that model performance reached a plateau at 25 chemicals; that is, adding more chemicals did not improve the classification in this data set.

One important conclusion from the investigations summarized in this section is that the gene signatures derived by any algorithm need to be checked for their mechanistic content. They should then only be used for classification if they represent functions and pathways relevant for the phenotype to be modeled. An earlier review of toxicogenomic investigations in the area of carcinogenicity characterization and prediction, based on several data sets available at that time [48], had come to the conclusion that gene expression profiles in such studies can reflect underlying modes or mechanisms of action and may therefore be useful in the prediction of chemical carcinogenicity in conjunction with conventional short-term genotoxicity studies. This conclusion again highlights the importance of mechanistic background in the profiles, which may also help to interpret the somewhat contrasting suggestions by different research groups of using either subacute (1–14-day) or subchronic (90-day) studies to derive gene expression profiles indicative of a potential for tumor induction in a certain organ. Both study durations may be reasonable, depending on what is to be modeled. To capture differences

between NGCs and GCs, earlier profiles may be more useful, whereas if one wants to model carcinogenic properties of compounds in general, gene expression profiles from somewhat longer duration studies may be more suitable, since by then the elicited changes should have been settled to a more robust profile comparable between carcinogens irrespective of primary class. It may be helpful to include a GC class when the goal is to specifically model NGCs, simply to differentiate NGCs from real GCs. This might have helped in case of the TG-GATEs approach where the model built with necrotic NGCs versus nonhepatocarcinogens also predicted GCs as necrotic NGCs. Thus, in any case, the compounds need to be well defined concerning the class of toxicants they should represent, and it could make sense to subdivide training set NGCs into classes with different primary modes of action, for example, necrosis/regenerative hyperplasia versus directly mitosis inducing. Furthermore, the number of different compounds per class should not be overly biased toward one versus the other class, and there should be a rationale for dose and time point selection. It may also generally be helpful to perform a mechanistic analysis in parallel.

Based on the various toxicogenomic investigations now published in conjunction with other developments in the regulatory area concerning carcinogenicity assessment in both the pharmaceutical and environmental/industrial chemical areas, the use of such analyses in these areas may be reevaluated. As described in Section 6.1.2, the ICH proposal for changes to the current carcinogenicity assessment of pharmaceutical suggests a flexible weight of evidence approach that considers all available data up to 6 (or 12)-month studies to decide on the value of actually performing a 2-year bioassay. Toxicogenomic analyses may be helpful here to clarify mechanisms and in this way potentially evaluate human relevance. In contrast, in case of environmental/industrial chemicals, where huge numbers of compounds are still untested, a predictive approach may be useful for prioritization of compounds to be tested in the 2-year bioassay for which high-sensitivity classification models would be desirable [43]. In this respect, Thomas *et al.* [43] made calculations about the requirements for a predictive approach in case of environmental/industrial chemicals. From NTP records, it can be derived that developing gene expression biomarkers for a combination of eight organs from rats and mice of specific sexes would identify nearly 80% of all positive chemicals in the rodent cancer bioassay. Based on their additional analysis that model performance reached a plateau at 25 chemicals, 8 (for the most important organs) times 25 (compounds per organ), that is, 200, chemical exposures may be required to develop these gene expression prediction models. If the same compound would induce tumors in several of these eight organs, even less studies may be needed. A further sector where short-term prediction models in conjunction with mechanistic evaluation could assist, for example, in compound ranking for safety evaluation may be in the agrochemical area, where the 2-year cancer bioassay will likely remain to be mandatory [5,49], since there are currently no discussions comparable to ICH S1.

6.2.3 Recent Developments: Transcriptional Benchmark Dose Modeling Based on Functional Analyses

As already alluded to above, classifier models should be built on mechanistically based signatures, otherwise they may not be robust enough to classify independent test compounds. A more recently developed approach also founded in mechanistic interpretation of compound-induced gene expression changes is to apply benchmark dose modeling on gene expression profiles at the single gene and functional category levels to aid in quantitative risk assessment [50,51]. Starting with gene expression profiles derived from samples after exposure of a biological model to at least three doses of a compound for a certain duration, genes showing a dose-dependent expression response are identified with, for example, ANOVA. The profiles of these genes are then fitted to a selection of standard statistical models, for example, linear or polynomial. With the model that best describes the data, one can estimate the benchmark dose at which the expression of the gene significantly deviates from that in control animals. After matching each gene with its corresponding functional category, for example, gene ontology (GO) categories, and calculation of summary values per category, a BMD level for the treatment-dependent change of a function can be derived [17]. Uncertainty factors, which are usually applied to "no observed adverse effect levels" (NOAELs), can also be applied to BMDs for the most sensitive category/function, to derive reference doses (RfDs) [52].

This approach was first applied and developed with gene expression data from nasal epithelium after a 6 h inhalation exposure of rats to four different concentrations of formaldehyde up to 15 ppm [51]. Formaldehyde is a major industrial chemical that in rodent models can induce regenerative cell proliferation and DNA–protein cross-links together with an increased incidence of nasal squamous cell carcinoma after longer term exposure with doses from 2 to 15 ppm. Controversy exists on the dose dependence of genotoxicity versus regenerative cell proliferation in case of formaldehyde-induced tumors (summarized in Ref. [51]). From the gene expression data in the above-mentioned study, average BMDs and benchmark dose lower confidence limits (BMDLs) were calculated for each GO category. Interestingly, the BMDs estimated for categories related to cell proliferation and DNA damage were similar to BMDs for induction of DNA–protein cross-links and DNA synthesis measured in prior labeling studies. Overall, the integration of BMD and gene expression analysis allowed to reveal dose-dependent effects of a compound at the molecular level and to derive RfDs for alterations of particular cellular processes. This is expected to greatly enhance chemical risk assessment. Thomas *et al.* continued to develop and expand that approach by applying it to further gene expression studies with lung and liver carcinogens in mice [52,53] and liver, bladder, and thyroid carcinogens in rats [13]. Based on these investigations, they finally developed a flowchart for the application of transcriptomic data to chemical risk assessment for both non-cancer- and cancer-related apical endpoints. Overall, this represents a pragmatic

approach to derive provisional point of departure (POD) dose values for chemicals without published RfDs. These studies also revealed that transcriptional perturbations do not occur at significantly lower doses than apical responses.

Such a BMD modeling for transcriptional responses was also applied to furan-induced gene expression changes in B6C3F1 mouse liver after 3 weeks of treatment with noncarcinogenic and carcinogenic doses [54]. The postulated cancer mode of action for furan includes regenerative hyperplasia following chronic cytotoxicity and inflammation, and a dose-response study for both 3 weeks and 2 years suggested the POD for carcinogenicity in B6C3F1 mice between 2 and 4 mg/kg [55]. The BMDLs calculated for hepatocellular carcinoma and adenoma as observed in the NTP mouse study, 0.92 and 1.57 mg/kg, respectively, were close to the suggested cancer POD, and the BMDLs for transcriptional responses representing pathways consistent with the postulated mode of action, for example, oxidative stress response, inflammation, and cell survival and growth, were also in this range [54]. This observed consistency between transcriptional responses and cancer data again demonstrated the applicability of toxicogenomics in quantitative risk assessment for environmental/industrial chemicals.

6.2.4 Recent Opportunities: Publicly Available Data

A prerequisite to perform toxicogenomic analysis is well-annotated gene expression data from sufficiently documented studies. Not only for predictive toxicogenomics, but also for mechanistic investigations and biomarker searching, it is helpful to have access to readily comparable data from the same biological model treated with various compounds. Two of the largest toxicogenomic databases were recently made public: TG-GATEs from the Toxicogenomics Project in Japan [56] and DrugMatrix, originally generated by Iconix Pharmaceuticals [44] (Table 6.2). Both databases contain gene expression profiles of mostly the liver after treatment of male Sprague Dawley rats and of hepatocytes *in vitro*, with compounds representing to a major part marketed drugs including several GCs and NGCs. Major advantages of the two databases compared with other publicly available data are (1) their uniform experimental design; (2) the availability of extensive metadata for classical toxicological endpoints; (3) the inclusion of a large number of profiled marketed drugs, which should allow an assessment of whether gene expression profiles measured in preclinical models can help predict human toxicity; and (4) the presence of *in vivo* and *in vitro* gene expression profiles after treatment with the same compounds, which should help to define the capabilities and limitations of *in vitro* systems. Apart from many commonalities between the two databases enabling meta-analyses across databases, for example, use of the same rat strain and an overlap in 73 compounds, two key differences need to be considered: (1) the microarray platform differs for a considerable part of the expression data, and (2) the determination of the MTD for dosing is different [12].

Table 6.2 Content of two published toxicogenomic databases, that is, TGP and DrugMatrix.

Data set	DrugMatrix (Iconix)	
Species	Male Sprague Dawley rat	
Gender	Male	
Study type	*In vivo*	*In vitro*
Dosing type	Repeat dose	Single dose
MTD definition	50% reduction in weight gain over 5 days	
Dose levels	Control, low (fully effective dose), high (= MTD)	Control and high
Sample collection after treatment	0.25, 1, 3, 5, plus some 7+ days	16 and 24 h
Biological replicates	Triplicate	Duplicate
Microarray platform	Affymetrix RG230_2.0 and CodeLink RU1 arrays	
No. of compounds tested in liver	CodeLink: 43 Affymetrix: 201	CodeLink:120 Affymetrix: 126
No. of arrays tested in liver	CodeLink: 5264 Affymetrix: 2218	CodeLink: 780 Affymetrix: 939
Microarray data available for other tissues	Kidney, heart, thigh muscle, bone marrow, spleen, brain, and intestine	Not available
Items examined	Histopathology, body/organ weight, food consumption, hematology, and blood chemistry	Not available

(continued)

Table 6.2 (*Continued*)

Data set	TGP (Japan)			
Species	Male Sprague Dawley rat	Male Sprague Dawley rat	Male Sprague Dawley rat	Human donor
Gender	Male			Not known
Study type	*In vivo*	*In vivo*	*In vitro*	*In vitro*
Dosing type	Single dose	Repeat dose	Single dose	Single dose
MTD definition	1-week dose range finding study → MTD			
Dose levels	Control, low, medium, high (= MTD) with a 1:3:10 ratio between dose levels	Control, low, medium, high (= MTD) with a 1:3:10 ratio between dose levels	Control, low, medium, high with a 1:5:25 ratio between dose levels	Control, low, medium, high with a 1:5:25 ratio between dose levels
Sample collection after treatment	3, 6, 9, and 24 h	24 h after treatment for 3, 7, 14, and 28 days	2, 8, and 24 h	2, 8, and 24 h
Biological replicates	Triplicate	Triplicate	Duplicate	Duplicate
Microarray platform	Affymetrix RG230_2.0 array	Affymetrix RG230_1.0 array	Affymetrix RG230_1.0 array	Affymetrix human U133 2.0 array
No. of compounds tested in liver	Phase I: 131 Phase II: 27	Phase I: 131 Phase II: 12	Phase I: 131 Phase II: 14	Phase I: 119 Phase II: 39
No. of arrays tested in liver	7378	6765	3370	2610
Microarray data available for other tissues	Kidney		Not available	Not available
Items examined	Histopathology, body/organ weight, food consumption, hematology, and blood chemistry	Histopathology, body/organ weight, food consumption, hematology, and blood chemistry	Cell viability (lactate dehydrogenase release and DNA)	

Modified from Ref. [12].
Abbreviations: Affx, Affymetrix; MTD, maximum tolerated dose.

Overall, the public availability of these comprehensive data is expected to stimulate renewed analysis and development of novel data mining tools leading to a more detailed insight into mechanisms of action. This may also enable the derivation of connections between genes and pathways relevant for the development of certain toxic phenotypes and may eventually allow distinction between adaptive and adverse effects at the molecular level. Since some of the compounds represent different mechanistic classes of carcinogens, these data may also be investigated to obtain further insight into early molecular effects of these compound types.

6.3 Conclusions and Outlook

The principal hypothesis underlying a toxicogenomic strategy is that chemical-specific patterns of altered gene expression can be revealed using high-density microarray analysis of tissues from exposed organisms. Analyses of these patterns should allow classification of toxicants and provide important mechanistic insights.

Many publications in this area have shown that gaining mechanistic insight is clearly possible, which will be further enhanced by application of newly developed analysis tools such as causal reasoning that infers upstream molecular events causing the observed gene expression changes [57–60]. Mechanistic toxicogenomics is now used in various research areas, for example, case by case for detailed analysis of observed phenotypes in drug development and for exploration of the mode(s) of action of environmentally important chemicals. In the latter area, toxicogenomic analyses may also help to prioritize compounds for further chronic testing, and together with benchmark dose modeling may even be used in risk assessment for derivation of reference doses.

Concerning its application as a predictive tool, its use is still rather selective. The reported studies in this area including those summarized above do highlight that for a predictive approach apart from a requirement of quite many data derived from uniformly performed studies, several other prerequisites need to be taken into account. One prerequisite appears to be that the selected signature genes represent functions and pathways likely involved in the endpoint to be modeled. The data available now in the public databases will allow revisiting several aspects in a more rigorous manner. It is clear from an investigation by the MAQC Consortium [61–63] that some endpoints can be modeled well whereas for others prediction accuracies will remain moderate.

References

1 Marone, P.A., Hall, W.C., and Hayes, A.W. (2013) Reassessing the two-year rodent carcinogenicity bioassay: a review of the applicability to human risk and current perspectives. *Regul. Toxicol. Pharmacol.*, **68**, 108–118.

2 ICH (1996) The Need for Carcinogenicity Studies of Pharmaceuticals. Guideline S1A. CPMP/ICH/140/95.

3 ICH (1998) Carcinogenicity: Testing for Carcinogenicity of Pharmaceuticals. Guideline S1B. CPMP/ICH/299/95.

4 ICH (2008) Dose selection for carcinogenicity studies. Guideline S1C(R2), Center for Drug Evaluation and Research; Center for Biologics Evaluation and Research: ICH revision 1. *Fed. Reg.*, **60**, 11278–11281. Addendum: **62**, 64260.

5 OECD (2009) Guideline for the Testing of Chemicals. Section 4 (Part 451): Health Effects: Carcinogenicity Studies.

6 Alden, C.L., Lynn, A., Bourdeau, A., Morton, D., Sistare, F.D., Kadambi, V.J., and Silverman, L. (2011) A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting for human risk. *Vet. Pathol.*, **48**, 772–784.

7 Reddy, M.V., Sistare, F.D., Christensen, J. S., Deluca, J.G., Wollenberg, G.K., and Degeorge, J.J. (2010) An evaluation of chronic 6- and 12-month rat toxicology studies as predictors of 2-year tumor outcome. *Vet. Pathol.*, **47**, 614–629.

8 Sistare, F.D., Morton, D., Alden, C., Christensen, J., Keller, D., Jonghe, S.D., Storer, R.D., Reddy, M.V., Kraynak, A., Trela, B. *et al.* (2011) An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. *Toxicol. Pathol.*, **39**, 716–744.

9 Morton, D., Bourcier, T., and Alden, C.L. (2013) Improving carcinogenicity assessment. *Toxicol. Pathol.*, **41**, 263–270.

10 ICH (2012) Concept Paper. S1: Rodent Carcinogenicity Studies for Human Pharmaceuticals. Available at www.ich.org/products/guidelines/safety/safety-single/article/new-topic-rodent-carcinogenicity-studies-for-human-pharmaceuticals.html (last accessed January 8, 2014).

11 FDA (2013) Department of Health and Human Services, ICH; proposed change to rodent carcinogenicity testing of pharmaceuticals; request for comments. Docket No. FDA-2013-N-0222. *Fed. Reg.*, **78**, 16681–16684.

12 Chen, M., Zhang, M., Borlak, J., and Tong, W. (2012) A decade of toxicogenomic research and its contribution to toxicological science. *Toxicol. Sci.*, **130**, 217–228.

13 Thomas, R.S., Wesselkamper, S.C., Wang, N.C., Zhao, Q.J., Petersen, D.D., Lambert, J.C., Cote, I., Yang, L., Healy, E., Black, M. B. *et al.* (2013) Temporal concordance between apical and transcriptional points of departure for chemical risk assessment. *Toxicol. Sci.*, **134**, 180–194.

14 Klaunig, J.E. (2013) Chemical carcinogenesis, in *Casarett and Doull's Toxicology: The Basic Science of Poisons* (ed. C.D. Klaassen), McGraw-Hill, pp. 393–443.

15 Osimitz, T.G., Droege, W., Boobis, A.R., and Lake, B.G. (2013) Evaluation of the utility of the lifetime mouse bioassay in the identification of cancer hazards for humans. *Food Chem. Toxicol.*, **60**, 550–562.

16 Afshari, C.A., Hamadeh, H.K., and Bushel, P.R. (2011) The evolution of bioinformatics in toxicology: advancing toxicogenomics. *Toxicol. Sci.*, **120** (Suppl. 1), S225–S237.

17 Yang, L., Allen, B.C., and Thomas, R.S. (2007) BMDExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics*, **8**, 387.

18 Oberemm, A., Ahr, H.J., Bannasch, P., Ellinger-Ziegelbauer, H., Gluckmann, M., Hellmann, J., Ittrich, C., Kopp-Schneider, A., Kramer, P.J., Krause, E. *et al.* (2009) Toxicogenomic analysis of *N*-nitrosomorpholine induced changes in rat liver: comparison of genomic and proteomic responses and anchoring to histopathological parameters. *Toxicol. Appl. Pharmacol.*, **241**, 230–245.

19 Oberemm, A., Ellinger-Ziegelbauer, H., Ahr, H.J., Kroeger, M., Weimer, M., Ittrich, C., Kopp-Schneider, A., Richter-Reichhelm, H.-B., and Gundert-Remy, U. (2007) Comparative analysis of *N*-nitrosomorpholine-induced differential gene and protein expression in rat liver. *Naunyn Schmiedebergs Arch. Pharmacol.*, **375**, 92.

20 Sato, K., Satoh, K., Tsuchida, S., Hatayama, I., Shen, H., Yokoyama, Y., Yamada, Y.,

and Tamai, K. (1992) Specific expression of glutathione *S*-transferase Pi forms in (pre)neoplastic tissues: their properties and functions. *Tohoku J. Exp. Med.*, **168**, 97–103.

21 Singh, G., Mallick, S., Sharma, V., Joshi, N., Purkait, S., Jha, P., Sharma, M.C., Suri, V., Julka, P.K., Mahapatra, A.K. *et al.* (2012) A study of clinico-pathological parameters and O^6-methylguanine DNA methyltransferase (MGMT) promoter methylation status in the prognostication of gliosarcoma. *Neuropathology*, **32**, 534–542.

22 Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H.J. (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term *in vivo* studies. *Mutat. Res.*, **637**, 23–39.

23 Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H.J. (2004) Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol. Sci.*, **77**, 19–34.

24 Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., and Ahr, H.J. (2005) Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. *Mutat. Res.*, **575**, 61–84.

25 Christmann, M. and Kaina, B. (2013) Transcriptional regulation of human DNA repair genes following genotoxic stress: trigger mechanisms, inducible responses and genotoxic adaptation. *Nucleic Acids Res.*, **41**, 8403–8420.

26 Ellinger-Ziegelbauer, H., Aubrecht, J., Kleinjans, J.C., and Ahr, H.J. (2009) Application of toxicogenomics to study mechanisms of genotoxicity and carcinogenicity. *Toxicol. Lett.*, **186**, 36–44.

27 Ellinger-Ziegelbauer, H., Fostel, J.M., Aruga, C., Bauer, D., Boitier, E., Deng, S., Dickinson, D., Le Fevre, A.C., Fornace, A. J., Jr., Grenet, O. *et al.* (2009) Characterization and interlaboratory comparison of a gene expression signature for differentiating genotoxic mechanisms. *Toxicol. Sci.*, **110**, 341–352.

28 Nakayama, K., Kawano, Y., Kawakami, Y., Moriwaki, N., Sekijima, M., Otsuka, M., Yakabe, Y., Miyaura, H., Saito, K., Sumida, K. *et al.* (2006) Differences in gene expression profiles in the liver between carcinogenic and non-carcinogenic isomers of compounds given to rats in a 28-day repeat-dose toxicity study. *Toxicol. Appl. Pharmacol.*, **217**, 299–307

29 Iida, M., Anna, C.H., Hartis, J., Bruno, M., Wetmore, B., Dubin, J.R., Sieber, S., Bennett, L., Cunningham, M.L., Paules, R. S. *et al.* (2003) Changes in global gene and protein expression during early mouse liver carcinogenesis induced by non-genotoxic model carcinogens oxazepam and Wyeth-14,643. *Carcinogenesis*, **24**, 757–770.

30 Iida, M., Anna, C.H., Holliday, W.M., Collins, J.B., Cunningham, M.L., Sills, R.C., and Devereux, T.R. (2005) Unique patterns of gene expression changes in liver after treatment of mice for 2 weeks with different known carcinogens and non-carcinogens. *Carcinogenesis*, **26**, 689–699.

31 Blanco-Bose, W.E., Murphy, M.J., Ehninger, A., Offner, S., Dubey, C., Huang, W., Moore, D.D., and Trumpp, A. (2008) c-Myc and its target FoxM1 are critical downstream effectors of constitutive androstane receptor (CAR) mediated direct liver hyperplasia. *Hepatology*, **48**, 1302–1311.

32 Chen, X., Muller, G.A., Quaas, M., Fischer, M., Han, N., Stutchbury, B., Sharrocks, A. D., and Engeland, K. (2013) The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. *Mol. Cell. Biol.*, **33**, 227–236.

33 Yap, C.S., Peterson, A.L., Castellani, G., Sedivy, J.M., and Neretti, N. (2011) Kinetic profiling of the c-Myc transcriptome and bioinformatic analysis of repressed gene promoters. *Cell Cycle*, **10**, 2184–2196.

34 Waring, J.F., Jolly, R.A., Ciurlionis, R., Lum, P.Y., Praestgaard, J.T., Morfitt, D.C., Buratto, B., Roberts, C., Schadt, E., and Ulrich, R.G. (2001) Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharmacol.*, **175**, 28–42.

35 Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant,

R., Stoll, R., Barrett, J.C., Paules, R.S. *et al.* (2002) Prediction of compound signature using high density gene expression profiling. *Toxicol. Sci.*, **67**, 232–240.
36 Steiner, G., Suter, L., Boess, F., Gasser, R., De Vera, M.C., Albertini, S., and Ruepp, S. (2004) Discriminating different classes of toxicants by transcript profiling. *Environ. Health Perspect.*, **112**, 1236–1248.
37 Nie, A.Y., McMillian, M., Brandon Parker, J., Leone, A., Bryant, S., Yieh, L., Bittner, A., Nelson, J., Carmen, A., Wan, J. *et al.* (2006) Predictive toxicogenomics approaches reveal underlying molecular mechanisms of nongenotoxic carcinogenicity. *Mol. Carcinog.*, **45**, 914–933.
38 Fielden, M.R., Brennan, R., and Gollub, J. (2007) A gene expression biomarker provides early prediction and mechanistic assessment of hepatic tumor induction by nongenotoxic chemicals. *Toxicol. Sci.*, **99**, 90–100.
39 Fielden, M.R., Nie, A., McMillian, M., Elangbam, C.S., Trela, B.A., Yang, Y., Dunn, R.T., 2nd, Dragan, Y., Fransson-Stehen, R., Bogdanffy, M. *et al.* (2008) Interlaboratory evaluation of genomic signatures for predicting carcinogenicity in the rat. *Toxicol. Sci.*, **103**, 28–34.
40 Fielden, M.R., Adai, A., Dunn, R.T., 2nd, Olaharski, A., Searfoss, G., Sina, J., Aubrecht, J., Boitier, E., Nioi, P., Auerbach, S. *et al.* (2011) Development and evaluation of a genomic signature for the prediction and mechanistic assessment of nongenotoxic hepatocarcinogens in the rat. *Toxicol. Sci.*, **124**, 54–74.
41 Auerbach, S.S., Shah, R.R., Mav, D., Smith, C.S., Walker, N.J., Vallant, M.K., Boorman, G.A., and Irwin, R.D. (2010) Predicting the hepatocarcinogenic potential of alkenylbenzene flavoring agents using toxicogenomics and machine learning. *Toxicol. Appl. Pharmacol.*, **243**, 300–314.
42 Uehara, T., Minowa, Y., Morikawa, Y., Kondo, C., Maruyama, T., Kato, I., Nakatsu, N., Igarashi, Y., Ono, A., Hayashi, H. *et al.* (2011) Prediction model of potential hepatocarcinogenicity of rat hepatocarcinogens using a large-scale toxicogenomics database. *Toxicol. Appl. Pharmacol.*, **255**, 297–306.
43 Thomas, R.S., Bao, W., Chu, T.M., Bessarabova, M., Nikolskaya, T., Nikolsky, Y., Andersen, M.E., and Wolfinger, R.D. (2009) Use of short-term transcriptional profiles to assess the long-term cancer-related safety of environmental and industrial chemicals. *Toxicol. Sci.*, **112**, 311–321.
44 Ganter, B., Tugendreich, S., Pearson, C.I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., Brady, L., Browne, L.J., Calvin, J.T., Day, G.J. *et al.* (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.*, **119**, 219–244.
45 Leopold, W.R., Miller, J.A., and Miller, E. C. (1979) *S*-Vinyl homocysteine, an analog of ethionine that is highly mutagenic for *S. typhimurium* TA100. *Biochem. Biophys. Res. Commun.*, **88**, 395–401.
46 National Research Council (NRC) (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*, The National Academies Press, Washington, DC.
47 Yang, H., Harrington, C.A., Vartanian, K., Coldren, C.D., Hall, R., and Churchill, G.A. (2008) Randomization in laboratory procedure is key to obtaining reproducible microarray results. *PLoS One*, **3**, e3724.
48 Waters, M.D., Jackson, M., and Lea, I. (2010) Characterizing and predicting carcinogenicity and mode of action using conventional and toxicogenomics methods. *Mutat. Res.*, **705**, 184–200.
49 EPA (2013) Data Requirements for Pesticide Registration. Available at http://www.epa.gov/opp00001/regulating/data_requirements.htm (last accessed January 8, 2013).
50 Andersen, M.E., Clewell, H.J., 3rd, Bermudez, E., Willson, G.A., and Thomas, R.S. (2008) Genomic signatures and dose-dependent transitions in nasal epithelial responses to inhaled formaldehyde in the rat. *Toxicol. Sci.*, **105**, 368–383.
51 Thomas, R.S., Allen, B.C., Nong, A., Yang, L., Bermudez, E., Clewell, H.J., 3rd, and Andersen, M.E. (2007) A method to integrate benchmark dose estimates with genomic data to assess the functional

effects of chemical exposure. *Toxicol. Sci.*, **98**, 240–248.

52 Thomas, R.S., Clewell, H.J., 3rd, Allen, B. C., Wesselkamper, S.C., Wang, N.C., Lambert, J.C., Hess-Wilson, J.K., Zhao, Q. J., and Andersen, M.E. (2011) Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol. Sci.*, **120**, 194–205.

53 Thomas, R.S., Clewell, H.J., 3rd, Allen, B. C., Yang, L., Healy, E., and Andersen, M.E. (2012) Integrating pathway-based transcriptomic data into quantitative chemical risk assessment: a five chemical case study. *Mutat. Res.*, **746**, 135–143.

54 Jackson, A.F., Williams, A., Recio, L., Waters, M.D., Lambert, I.B., and Yauk, C. L. (2014) Case study on the utility of hepatic global gene expression profiling in the risk assessment of the carcinogen furan. *Toxicol. Appl. Pharmacol.*, **274**, 63–77.

55 Moser, G.J., Foley, J., Burnett, M., Goldsworthy, T.L., and Maronpot, R. (2009) Furan-induced dose–response relationships for liver cytotoxicity, cell proliferation, and tumorigenicity (furan-induced liver tumorigenicity). *Exp. Toxicol. Pathol.*, **61**, 101–111.

56 Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., and Urushidani, T. (2010) The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.*, **54**, 218–227.

57 Chindelevitch, L., Ziemek, D., Enayetallah, A., Randhawa, R., Sidders, B., Brockel, C., and Huang, E.S. (2012) Causal reasoning on biological networks: interpreting transcriptional changes. *Bioinformatics*, **28**, 1114–1121.

58 Enayetallah, A.E., Puppala, D., Ziemek, D., Fischer, J.E., Kantesaria, S., and Pletcher, M.T. (2013) Assessing the translatability of *in vivo* cardiotoxicity mechanisms to *in vitro* models using causal reasoning. *BMC Pharmacol. Toxicol.*, **14**, 46.

59 Enayetallah, A.E., Ziemek, D., Leininger, M.T., Randhawa, R., Yang, J., Manion, T. B., Mather, D.E., Zavadoski, W.J., Kuhn, M., Treadway, J.L. *et al.* (2011) Modeling the mechanism of action of a DGAT1 inhibitor using a causal reasoning platform. *PLoS One*, **6**, e27009.

60 Laifenfeld, D., Qiu, L., Swiss, R., Park, J., Macoritto, M., Will, Y., Younis, H.S., and Lawton, M. (2014) Utilization of causal reasoning of hepatic gene expression in rats to identify molecular pathways of idiosyncratic drug-induced liver injury. *Toxicol. Sci.*, **137**, 234–248.

61 Shi, L., Campbell, G., Jones, W.D., Campagne, F., Wen, Z., Walker, S.J., Su, Z., Chu, T.M., Goodsaid, F.M., Pusztai, L. *et al.* (2010) The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.*, **28**, 827–838.

62 Shi, W., Bessarabova, M., Dosymbekov, D., Dezso, Z., Nikolskaya, T., Dudoladova, M., Serebryiskaya, T., Bugrim, A., Guryanov, A., Brennan, R.J. *et al.* (2010) Functional analysis of multiple genomic signatures demonstrates that classification algorithms choose phenotype-related genes. *Pharmacogenomics J.*, **10**, 310–323.

63 Tillinghast, G.W. (2010) Microarrays in the clinic. *Nat. Biotechnol.*, **28**, 810–812.

7
Discovery and Application of Novel Biomarkers

Timothy W. Gant, Emma L. Marczylo, and Martin O. Leonard

7.1
Introduction

Biomarkers can be defined in various ways. In pathophysiology, a biomarker is a molecule that is an indicator for tissue damage that cannot be measured directly, or quantitated accurately. In this case, the molecule measured acts as a surrogate for the actual process of interest that cannot be measured directly, a latent variable. The US National Institutes of Health (NIH) defines a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention." New technologies, including high-throughput omics, have provided novel high-throughput methods for rapidly discovering and assessing changes in DNA (genetics and epigenetics), RNA (mRNA and miRNA), metabolites, proteins, and lipids. Some of these changes can also be biomarkers of past exposure or prognostic for future events. As such, there is a need to expand this biomarker definition beyond the present to include the biological measures indicative of something that has happened in the past or will happen in the future, for example, exposure to an environmental chemical that has been cleared from the body but has left behind a mark of its presence on the cellular macromolecules such as DNA and proteins. This is analogous to the fossil of a soft-bodied organism that lived long ago. The body of the organism has decayed, but a mark has been left in the rock that is indicative of its presence. In this scenario, the measured parameter is a marker of a past exposure or event. Are such biomarkers a realistic proposition? The potential has yet to be fully realized, but the science of epigenetics has revealed one example. Throughout life our DNA is epigenetically modified to induce changes in gene expression that enable us to adapt to our environment. Such epigenetic marks have the potential to act as indicators of past exposure. Other novel biomarkers can be prognostic biomarkers for future events. These are perhaps among the most valuable biomarkers because they allow the possibility of intervention before the adverse event occurs, or avoidance altogether. For example, a good prognostic biomarker of liver failure outcome from acetaminophen overdose would allow

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

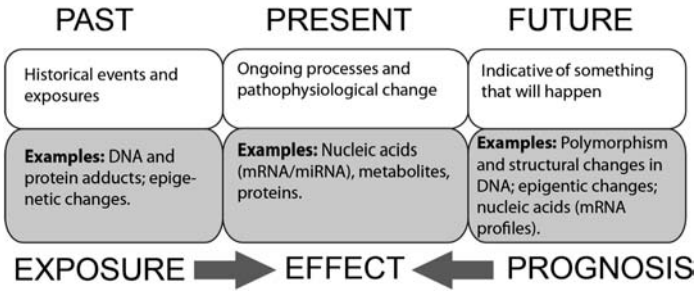


Figure 7.1 Division of biomarker classes into past, present, and future.

prediction of those patients where liver failure was likely to occur and thus identify patients requiring aggressive treatment with *N*-acetylcysteine to try and alter the development or severity of the liver failure. Similarly, a gene polymorphism may have no adverse consequence for the carrier under normal conditions, but may be adverse after exposure to a drug or chemical. In this scenario, assessment of the polymorphic marker would allow active steps to be taken to avoid exposure and thus prevent an adverse outcome. Later in this chapter we will return to specific examples in all of these scenarios utilizing, in the main, genomics (analysis of events at the gene level), although many of the points considered are also applicable to proteomics and metabolomics.

In summary, biomarkers can be divided into three time-based classes: past, present, and future. Past markers are exposure biomarkers and indicate that an exposure has taken place even if there is no adverse consequence from that exposure. Present markers are those that indicate an ongoing process, such as pathophysiology. Future markers are prognostic. They indicate something that is likely to happen but has not yet happened (Figure 7.1).

For all the above classes, there are common properties that are required for a successful biomarker. These are sensitivity, specificity, accessibility, ease of assay, and, for a biomarker that is a surrogate for the latent variable of interest, a linear relationship to the latent variable.

7.1.1 New Technologies Give Rise to Novel Opportunities for Biomarker Discovery

Over the past 15 years, omics technologies have been developed. There are various types, but they have one aspect in common, high-throughput analysis. Such methods started with the development of automated capillary sequencing that enabled the relatively rapid sequencing of mRNA libraries, providing access to many more gene probes than had been available previously. The techniques of the time that utilized such probes to measure gene expression, primarily the Northern and dot blot, did not have the capacity to permit the high-throughput usage of these large numbers of new gene probes. A new method had to be found, and it was the microarray. In essence, microarrays are simply a solid

hybridization surface on which a large number of probes can be deposited, for the first time allowing the simultaneous assay of many different genes in the same experiment. Further developments in high-throughput sequencing (HTS), where many millions of DNA bases can be sequenced in a single run, have now largely replaced microarrays. However, the underlying principle remains the same: the simultaneous determination of the expression of many genes within one sample. Indeed, the capacity of the HTS machines has now increased to such an extent that whole genomes can be sequenced in a single run. Consequently, the cost per human genome has fallen from $300 000 000 in 2003 for the sequencing of the first human genome to now a claimed $1000 per genome (http://www.illumina.com/systems/hiseq-x-sequencing-system.ilmn). Such advances have enabled the high-throughput study of a wide range of genome and epigenome changes in addition to gene expression, including polymorphisms and alterations in DNA methylation and histone modifications. Furthermore, it is not just analysis at the nucleic acid level that has reached the high-throughput level by the introduction of new technology. Developments in mass spectrometry analysis and nuclear magnetic resonance spectroscopy have had a similar effect on the analysis of proteins and metabolites. Together these technologies have given rise to three major branches of omics screening: genomics, proteomics, and metabolomics (nucleic acids, proteins, and metabolites, respectively). This has provided a wealth of opportunities for the discovery and validation of novel biomarkers, the subject of this chapter.

7.2 Novel RNA Biomarkers

7.2.1 The Complex RNA Biomarker in Cancer

One of the first applications of the new genomic technologies was in the development of the complex RNA biomarker (transcriptomic biomarker), particularly for distinguishing between various types of cancer that were difficult to differentiate pathologically. This was first described by Golub *et al.* in a landmark paper in *Science* in 1999 [1]. Here the authors showed for the first time how a complex gene expression profile could be used to separate acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) without any prior knowledge of the tumor type. This approach was subsequently applied to solid tumors, including diffuse large B-cell lymphoma, where complex gene expression profiles were able to separate four distinct subtypes of tumor that could not be identified by other means [2,3], and carcinoma of the bladder where gene expression profiles were able to identify three major stages and further subgroups of one of these stages [4]. Such transcriptional profiling work has since been extended to a wide range of other tumor types, further demonstrating how complex transcriptomic

biomarkers of gene expression can be used to distinguish between closely related pathologies.

Use of the transcriptomic biomarker for the pathological identification of tumors would fit within the NIH biomarker definition as it is a marker of an ongoing process, or within the terminology introduced here, a present biomarker. However, transcriptomic biomarkers have also been shown to have utility as prognostic or future biomarkers. One of the earliest and best examples of a prognostic transcriptomic biomarker was identified in a breast cancer study carried out by van't Veer *et al.* in 2002 [5]. From 78 primary breast cancers, a gene expression profile of 5000 differentially expressed genes was developed. From these 78 patients, 34 developed distant metastases within 5 years of diagnosis, while 44 continued to be disease free. The gene expression profile of 5000 differentially expressed genes was subject to a supervised clustering, which identified a gene expression profile of 70 genes that could separate the tumors with good prognosis (no metastases at 5 years) from those with poorer prognosis (metastases within 5 years). This gene signature was validated over a number of years, eventually resulting in the diagnostic profile named the MammaPrint® [6]. The MammaPrint gene set has shown validity in predicting both the time of remaining free from distant metastases and the time of overall survival, and is now licensed as a valid diagnostic test using microarrays as the testing platform. Many other such studies have also been performed, for example, in nervous system embryonal tumors [7], but none have yet been transferred from the laboratory to the clinic as effectively as the MammaPrint.

The potential utility of such transcriptomic biomarkers to not only predict a current state, such as pathology, but also act as a prognostic marker for future outcome is in many ways quite remarkable. Solid tumors such as breast carcinomas and others are not a single cell type but consist of a multitude of cancer and noncancer cell types. All of the noncancer cell types within the tumor, such as blood vessels, stroma, and inflammatory cells, will dilute the overall gene signature. Yet, a gene signature indicative of the phenotypic characteristics of the tumor can survive the dilution effect and still be adequately detected.

Another remarkable aspect of this type of complex gene expression signature is that it works even though many genes are dynamically regulated and respond to the individual environments in which they are located. The fact that a set of genes such as the 70 genes corresponding to the MammaPrint are commonly regulated across a set of tumors and contribute to a specific outcome suggests that they are differentially expressed as a result of a change in the underlying tumor genotype that is relatively stable. One such genome change is gene duplication and deletion. These are structural alterations that commonly occur in the DNA of tumor cells. Gene duplication and deletion events can thus also act as novel present and future biomarkers, and are easily detected using microarray-based comparative genome hybridization.

A disadvantage of RNA-based gene expression methods is their dependence on a reference. Levels of gene expression at the RNA level cannot be analyzed alone and interpreted by comparison with a historic reference level in the same

manner as, for example, enzyme activity. RNA can only be measured by reference to something else. The MammaPrint depends on a standard reference. Standard references can be difficult to reproducibly manufacture and may not be standard across laboratories, raising the possibility that a gene expression profile generated in one laboratory may not match that generated in another laboratory. For these reasons, the MammaPrint assay is undertaken in one specialist laboratory [8]. Furthermore, RNA is sensitive to degradation through contamination with environmental RNases. These drawbacks have the very real possibility of limiting the widespread application of such biomarkers to just a few specific testing laboratories. Another disadvantage of this method is that it requires a sample of tissue. While this may be accessible at initial diagnosis, it makes following the tumor, particularly during treatment, challenging.

7.2.2 The Complex RNA Biomarker in Toxicology

The same approach as that used in oncology has been used to develop transcriptional biomarkers of exposure to toxins, adverse outcomes, and pathological changes in the field of toxicology. Several consortia have been involved in such work, one example of which is the FP6 PredTox [9], although there are others. Most of these studies have been with *in vivo* animal or *in vitro* models and therefore have the advantage of a control sample for comparison. These approaches can, without doubt, identify gene expression profiles indicative of events. However, the analysis of results has proven more difficult than that for cancer. Commonly, the gene expression profile has been identified from tissues where there is already a pathological change underway as a result of the chemical insult. This can make it difficult to discern those gene expressions that are due to the toxicodynamics of the chemical from those due to the pathological change in the organ. One of the most popular early analysis methods of such genomic data was so-called phenotypic anchoring. Here the definition of phenotypic anchoring is the linkage of gene expression change(s) with a current pathological change. It is also possible to link gene expression change(s) with a future outcome, an analysis method better referred to as prognostic anchoring. The use of phenotypic anchoring is demonstrated in the early work of Moggs *et al.* [10]. Work such as this is elegant and indicates the close relationship between specific pathology and gene expression in a similar manner to that found in cancer. However, a difficult question has to be asked: What do data such as these actually tell us? The gene expression profile may inform on the mechanism by which the pathology occurs and may be able to differentiate pathologies not easily distinguished by histopathology, but it does not generally predict the occurrence of a toxic event before it actually takes place. This is not a criticism of the early work. Linking the data to a pathological change did provide a means to make some sense of the large genomic data sets generated. Indeed, we have performed similar analyses ourselves [11,12], as have many others. If, however, we return to the study of Moggs *et al.* [10], it could be argued that the blotted uterine weight was

a more sensitive biomarker of the estrogenic effect than the gene expression profile. Phenotypically anchored studies are therefore scientifically interesting, can inform on pathology, and provide a greater insight into that pathology than histology alone, but usually do not provide a transcriptomic biomarker of a toxic event that cannot be discerned by another analysis method in the tissue itself. Where they do have utility though is in the identification of changes that may be detected in biofluids distant from the pathophysiological change. This then provides an opportunity for finding a biomarker that can be measured away from the site of toxicity, which indicates the ongoing process without a need to sample the tissue. These biomarkers are valuable because they allow a continuous monitoring of the processes.

While gene expression analysis offers advantages to histopathology [13], the limitation of the transcriptomic biomarker is again tissue accessibility. Consider, for example, oncology. As previously mentioned, at first diagnosis biopsy or surgically removed tumor may be available. Such tissue samples are unlikely to be available subsequently if the tumor progresses, precisely the time when a biomarker would be particularly useful for monitoring the process that cannot be measured directly, tumor progression and development. The same is true in toxicology where it is even less likely that there will be tissue available from the damaged organ, particularly in cases of human toxicity. A transcriptomic biomarker of tissue damage in a readily accessible tissue is therefore of even more importance. Intuitively, it might be anticipated that gene expression changes giving rise to a transcriptomic biomarker would be limited to the organ in which the damage was located. However, it was elegantly demonstrated by Bushel *et al.* that this is not necessarily the case [14]. Here the authors monitored mRNAs in the blood of rats undergoing acetaminophen-induced liver toxicity, comparing the predictive power of four gene expression models with that of three independent pathologists, clinical chemistry and hematology. In each case, the mRNA profiles were found to outperform the more traditional biomarkers.

To conclude, the transcriptional biomarker has utility in the pathological analysis and quantitation of that pathology, can inform on pathological mechanisms, and may be linked to an outcome, such as propensity to metastasize. In this respect, it has the potential to act as a prognostic biomarker. Its analytical complexity, however, can be limiting, and consideration has to be given as to whether it truly outperforms the traditional biomarkers, particularly the cellular change that is ongoing at the time of tissue sampling.

7.2.3
Connectivity Mapping with the Complex RNA Biomarker for Hazard Identification

Connectivity mapping is a method first published by Lamb *et al.* in 2006 [15] and further adapted by others [16,17]. The concept is very simple. It provides a method to link a gene expression profile with others in a database. As such, it requires a gene expression signature of interest, a database of other gene expression signatures that are linked to other metadata such as chemical name/type,

pathological change, tissue or cell type, and an algorithm to link the gene expression signature of interest with those in the database. The database generated by the Broad Institute used four cell lines treated with 1309 chemicals largely, although not exclusively, at one dose level. There are obvious limitations to this approach, in particular the lack of dose response. Nevertheless, the method was shown to be effective, initially for recognizing estrogen receptor agonists using a gene expression signature from cells treated with estrogen. Perhaps more interestingly, the method was also able to easily detect estrogen receptor antagonists by utilizing the exactly opposite gene expression signature [15,17]. The estrogen receptor is, however, a transcription factor in its own right with a specific DNA binding site. Therefore, it would be anticipated to create quite a specific gene expression biomarker that would have utility in an approach such as connectivity mapping. What about a target less specific for transcriptional gene regulation than a nuclear receptor; will the method still work? Recently, Caiment *et al.* have shown that mutagens and carcinogens, with no defined receptor, can also be identified using this approach [18]. The database can also be used in another way. By grouping chemicals that cause similar changes in gene expression, common patterns or signatures may be recognized that are indicative of mechanisms. For example, all mutagens in the database may be grouped and then by looking for genes common to the group that are differentially expressed compared with other groups a transcription biomarker may be identified that is indicative of mutagens.

7.2.4
miRNA Biomarkers

First discovered in *Caenorhabditis elegans* in 1993 [19], miRNAs are 21–25-nucleotide nonprotein-coding RNA species that are transcribed from non-protein-coding regions of the genome under the same control mechanisms as protein-coding genes that utilize RNA polymerase II [20]. They are primarily located in separate parts of the genome under their own transcriptional promoter, in introns, or less commonly in exons (Table 7.1). In the latter two cases, they are under the control of the promoter for that protein-coding gene. In all cases, they are spliced from a longer primary miRNA transcript by the RNase enzyme DROSHA and then exported as precursor miRNAs from the nucleus by the active transport pump EXPORTIN5. They are then further processed by the

Table 7.1 Genomic location of miRNAs in four species.

Genomic location	Human	Mouse	Rat	Dog
Intergenic	396 (52%)	405 (62%)	359 (76%)	225 (64%)
Intronic	308 (40%)	235 (24%)	112 (24%)	107 (31%)
Exonic	58 (8%)	12 (2%)	0 (0%)	18 (5%)

From Ref. [21].

RNase enzyme DICER before being incorporated onto the RISC complex from which they modulate translation by either suppressing translation or leading to cleavage and thus degradation of their target mRNA strand [20].

miRNAs are important in toxicology because they control rapid protein translation from mRNA strands and can therefore be utilized to provide a very rapid change in protein levels in response to cellular stress [22]. They can also act as biomarkers in one of the two ways: first in the tissue in the same manner as mRNA because their transcription is regulated and can be modulated in response to cellular stress, and second as biomarkers in tissue fluids.

Remarkably, miRNAs are found in all body fluids (Table 7.2), are relatively stable to degradation, and are comparatively easily assayed. This has led to an interest in them both as biomarkers of toxicity within the target organ and as distant markers in an accessible body fluid. One of the first studies to exploit these properties was that published by Wang *et al.* in 2009, which used the established acetaminophen model of liver injury in rats and microarrays to assess differential levels of miRNAs in the plasma of these animals [23]. A number of miRNAs were shown to be altered, several of which demonstrated substantial changes up to approximately 450-fold over control. The two plasma miRNAs with the greatest increase in concentration were miRNA-122 and miRNA-192. Both of these miRNAs are known to be highly expressed in the liver, therefore supporting a hypothesis that these are released into the circulation as a result of liver damage [24]. These workers also performed a time and dose response for both of these miRNA species using four doses of acetaminophen from 0 to 300 mg/kg and two time points of 1 and 3 h postexposure. These analyses showed that both miRNA species were more sensitive than the traditional biomarker ALT for the detection of acetaminophen-induced liver damage. A similar analysis has been carried out in human subjects suffering from acute liver injury (ALI) [25]. Circulating miRNA-122 was significantly elevated from baseline in

Table 7.2 Detectable miRNAs in bodily fluids.

Body fluid	Number of detectable miRNAs
Amniotic fluid	359
Breast milk	429
Bronchial lavage	260
CSF	212
Colostrum	386
Peritoneal fluid	397
Plasma	349
Pleural fluid	210
Saliva	458
Seminal fluid	436
Tears	320
Urine	204

From Ref. [26].

patients suffering from both acetaminophen-induced and non-acetaminophen-induced ALI, while circulating miRNA-192 was significantly elevated only in patients suffering from acetaminophen-induced ALI. This suggests that miRNA-122 was indicative of general liver damage, while miRNA-192 was specific to acetaminophen-induced liver injury. Thus, like transcriptomic biomarkers, miRNA expression profiles rather than single miRNAs may prove to be most useful as biomarkers of specific pathologies.

Cardiovascular development has been one of the most investigated physiological systems for determining the role of miRNAs in physiology, and multiple studies have revealed that miRNAs are critical to many areas of this process [27]. Not surprisingly, therefore, the utility of miRNAs as biomarkers of cardiovascular disease has been extensively investigated. One of the first studies was that of Fichtlscherer *et al.*, which used a microarray profiling method to identify candidate miRNAs altered in the plasma of patients suffering from cardiovascular disease, and then further analyzed these candidate miRNAs with a quantitative real-time PCR assay on a hypothesis testing basis in a validation patient cohort [28]. Most of the miRNAs identified were decreased rather than increased in expression during cardiovascular disease.

Downregulation of miRNA species in plasma as a biomarker of a pathophysiological event is something that requires some additional thought and discussion. There are two aspects: first a downregulated miRNA runs counter to the expectation that a biomarker will be released from a tissue on damage, and second the measurement space (see below) that can be assigned to this event. Fichtlscherer *et al.* had this same thought and offered the hypothesis that the miRNA was being lost to the plasma because of uptake into atherosclerotic lesions [28]. However, they provided no experimental evidence in support of this hypothesis. Even if this is a reasonable explanation, it still renders downregulated miRNA biomarkers difficult to deal with analytically. There are two reasons for this: first they become progressively more difficult to measure as they decrease in abundance and are therefore of less value, and second their range for change is limited to between 1 (unchanged) and 0 (completely lost). In practice, limits of detection would most likely reduce this range even further. In contrast, an upregulated miRNA is not only easier to envisage hypothetically as being due to release from damaged tissue, but also progressively easier to measure and has an unlimited range for change from 1 to infinity. Therefore, given a choice between a biomarker that is upregulated upon tissue damage and one that is downregulated, the biomarker that is increased in expression offers many more advantages.

Returning to the work of Starkey Lewis *et al.* investigating miRNA biomarkers for the detection of acetaminophen-induced ALI, the data also showed that the candidate miRNA biomarkers were neither more sensitive nor more specific than the traditional biomarker ALT [25]. The question that then obviously arises is: Do these miRNA biomarkers have greater utility than the ALT marker that is well established for liver injury? From these data, the answer at the moment would have to be no. While this appears a harsh judgment on the science, it is important to consider that this field is relatively early in the research process.

Most of the data generated have been in animal models, so a great deal of work needs to be done to establish the utility, or lack thereof, of these miRNA biomarkers in the diagnosis of clinical disease and toxicology. Such work is actively ongoing and it is beyond the scope of this chapter. However, a literature search indicates that serum miRNAs are of interest in many disease types and in normal life processes such as aging and pregnancy. It is therefore likely that over time new candidate miRNA biomarkers will emerge and be validated, which will either have a greater utility than existing biomarkers or fulfill a need where there is no current biomarker.

7.3 DNA as a Biomarker

There are potentially three forms of DNA alteration that can act as biomarkers of toxicological events. The first of these is a future prognostic biomarker. This will take the form of some alteration in the genome leading to an alteration of gene function or expression that has a phenotypic outcome in the future. For example, a DNA polymorphism can be predictive of a future disease state such as breast cancer or of susceptibility or resistance to an environmental exposure. The second and third are biomarkers of exposure, so indicative of something that has happened in the past. These are DNA adducts and epigenetic changes. Both of these DNA alterations indicate that an exposure has taken place. They may also be indicative of a future adverse outcome, though not necessarily so.

The advent of high-throughput sequencing has led to a substantial increase in the number of known DNA polymorphisms and other DNA alterations such as deletions and duplications. These data are increasing at a substantial rate due to genetic variant analysis projects, such as HapMap [29] and 1000 Genomes (http://www.wellcome.ac.uk/Funding/Biomedical-science/Funded-projects/Major-initiatives/WTDV029748.htm), in addition to individual efforts estimated to have generated 30 000 human genomes by the end of 2011 [30]. These projects are starting to give some indication of the genetic biodiversity in the human population. As this diversity becomes associated with phenotypes, it can act as prognostic biomarkers of future events. For example, a number of genetic polymorphisms such as *BRCA1* mutations are prognostic for the development of cancer [31].

7.3.1 DNA Polymorphisms as Future Biomarkers of Disease and Xenobiotic Susceptibility

Many genes are involved in the response of cells to xenobiotics, including initial interaction(s), metabolism and excretion, response to cellular stress, repair of damage, or as a last resort induction of cell death processes. There is therefore substantial opportunity for variation in the susceptibility of individuals or subpopulations to particular xenobiotics as a result of genetic polymorphisms that

cause alterations in either the activity or abundance of the proteins involved in these processes.

For example, polymorphisms in the coding sequences of genes associated with xenobiotic metabolism can affect the manner in which these xenobiotics are metabolized. This has been demonstrated for various genes and compounds in an extensive number of publications. Johansson and Ingelman-Sundberg have reviewed some of these examples, including warfarin, where a knowledge of DNA polymorphisms in specific cytochrome p450 genes can be used to predict those that might have adverse sensitivity [32]. It is not just the metabolic genes that are important in determining sensitivity. Once metabolized to hydrophilic derivatives, metabolites often have to be moved across membranes and this can require active transport. One set of proteins responsible for this are the ATP binding cassette genes that utilize ATP to actively excrete such metabolites. This diverse family of proteins has many known polymorphisms, some of which lead to disease phenotypes and others to alteration in susceptibility to chemicals and drugs [33].

Polymorphic gene variations that confer susceptibility are not restricted to a single base sequence change, but also encompass structural variations such as copy number. Changes in the copy number (including both insertion and deletion) of genomic regions are potentially of more significance for conferring phenotype and differential sensitivity to xenobiotics than DNA polymorphisms. While polymorphisms affect the function and expression of one gene, copy number variations can affect whole stretches of DNA and the expression of many genes within that region. Copy number variation is also very common in both humans and experimental animals, although the full landscape of the variation has still to be explored in both [34,35].

These are just two examples of many DNA alterations that could be cited to indicate the ability of DNA to act as a biomarker of susceptibility to xenobiotic exposure. The purpose here is not to provide an extensive review but rather to highlight the area. The developments in DNA sequencing will reveal a great many more polymorphisms and structural variations in DNA. A proportion of these will not have relevance for either disease or toxicology, but others will. Being able to recognize these pathologically relevant alterations within the background data will be challenging, particularly as these variations might be rare, affecting subpopulations not represented in model species and test systems. Arguably, the only way forward is to develop a much greater understanding of specific mechanisms of toxicity, including recognition of the important genes associated with these mechanisms – and the metabolism and excretion of the xenobiotic. Large-scale genome projects will provide a rich source of information about genetic variation in the population, such as frequency of specific variations and identification of rare variants. Similarly, data will be available on the genetic differences between humans and model species. With knowledge of mechanisms of action, these data can be utilized to make predictions of susceptibility or resistance and thus genomic data will act as a future prognostic biomarker of differential effect(s).

7.3.2 DNA and Protein Adduct Biomarkers

DNA adducts are formed after activation of a xenobiotic to an electrophile that reacts with nucleophilic sites on DNA, of which the most common is the N7 of the guanine base [36]. Depending on the nature and site of the DNA adduct, it either is repaired or leads to mutagenesis [37]. There are many examples of such chemicals and among the "classics" are benzo[*a*]pyrene, 2-acetylamino fluorine, and the estrogen receptor antagonist drug tamoxifen [38–40]. As these adducts are removed and repaired from the DNA with a half-life that is tissue and type dependent (see Refs [41,42] as two examples of many), they can be measured and act as historical biomarkers of past exposure. Indeed, fetal DNA adducts induced by benzo[*a*]pyrene in the fetus can act as a biomarker of maternal smoking [43]. A similar process can be used with adducts formed on proteins [44]. The use of these DNA and protein adducts as biomarkers of past exposure has been greatly facilitated by the recent rapid developments in mass spectrometry methods, which now enable low parts per billion resolution and sensitivity in adduct detection.

7.3.3 Epigenetic Biomarkers

Epigenetic marks are chemical changes to either DNA or the histone proteins around which DNA is packaged that alter gene expression without a change in DNA sequence. The most abundant epigenetic mark on DNA is 5′-methylcytosine (5mC). This mark, or one of its derivatives, is found at a frequency of approximately 1–2% across the genome. In general, 5mC is a repressive mark that, when present within the promoter or enhancer regions of genes, leads to the reduced expression of that gene [45]. 5mC is recognized by the MECP2 protein, which directly binds 5mC and suppresses transcription. 5mC is initially deactivated by conversion to 5′-hydroxyl methylcytosine (5hmC). 5hmC is not recognized at DNA replication by the DNMT enzymes responsible for methylating DNA, and thus ultimately leads to loss of the DNA methylation mark. In contrast, 5mC can be transferred through mitotic and meiotic divisions, and is thus passed from mother cell to daughter cell in somatic tissues and also from parent to child via the gamete. Therefore, changes in methylation marks due to chemical or drug exposure can be transferred across cell and organism generations, and are thought to be the mechanism by which adaptive DNA changes can be passed from parent to child. This mechanism is thought to be behind the evolutionary inheritance of acquired traits proposed by Jean-Baptiste Lamarck. The modern epigenetic Lamarckian inheritance proposes that a beneficial phenotypic adaptation to the environment could be captured through the epigenetic modification of gene expression and passed onto the offspring via the germline so that any resulting children also benefit from the adaptation. A more traditional genetic mechanism involving a change in the DNA sequence itself would be too slow to enable such adaptive inheritance across a single generation. Such adaptive epigenetic inheritance has

been demonstrated in a rat model of carbon tetrachloride-induced fibrotic liver damage, where the progeny of rats with a history of such liver damage showed corresponding epigenetic and transcriptomic alterations in fibrosis-related genes and were more resistant to fibrogenesis when exposed to the same carbon tetrachloride regimen [46]. Such epigenetic processes could also enable the inheritance of detrimental phenotypes, and are thought to explain the transmission of some of the adverse disease outcomes described by numerous independent research groups in response to a wide range of environmental exposures. Indeed, in rats poor maternal care in early postnatal life has been associated with heritable epigenetic modifications in neuronal genes, altered neuronal gene expression, and adverse behavioral responses [47]. As a result of the stability of DNA methylation marks across multiple generations, they have been postulated to be potentially useful as biomarkers of past exposures. Perhaps even more interestingly, the putative involvement of such marks in the mechanism(s) of phenotypic inheritance may lead to the development of novel epigenetic biomarkers of present effect(s) and/or future pathologies.

It needs to be noted here that the ideas of Jean-Baptiste Lamarck have not been fully embraced by the scientific community and there is an active debate with respect to their significance in relation to multigenerational toxicity. This is explored in reviews by Youngson and Whitelaw [48] and Morgan *et al.* [49]. In both reviews, the authors make the valid point that there is a short period in early development when the primordial germ cells, the cells that give rise to the gametes, are formed and undergo extensive global epigenetic reprogramming. This means that there may be a limited time window during which chemicals and drugs could influence the removal and reestablishment, and thus inheritance, of such epigenetic marks. This perhaps explains why the majority of multigenerational effects assigned to these drugs and chemicals to date have utilized *in utero* exposure during the first trimester [50]. However, the studies in fibrotic response to liver damage and maternal care, where the adaptation was passed through generations without *in utero* exposure, suggest that there are wider mechanisms involving more than just a short vulnerable window in *in utero* exposure at play [46,47].

Histones carry modifications on the tails that extend from the core of the histone protein. There are five common modifications (acetylation, sumoylation, methylation, ubiquitination, and phosphorylation), of which the most common are methylation and acetylation (for a review see Ref. [51]). These marks can be added and removed by relatively dynamic processes in comparison with DNA methylation marks. This makes them of interest in mechanisms of toxicity, but arguably less suitable as biomarkers. They will therefore not be considered further in this chapter.

In 2008, ECETOC held a workshop that considered the use of epigenetic events as biomarkers of exposure. This workshop was very much a horizon scanning activity as good studies were not available at the time, and the mechanisms and consequences of such epigenetic events were still being established (ECETOC Workshop Report 09, 2009). Nevertheless, this workshop did serve to raise interest in the area, which is now starting to be realized.

Can epigenetic change act as a historical biomarker of past exposure? In the absence of a complete knowledge about the control mechanisms that lead to the establishment of epigenetic marks, one place to start is the biochemistry of cytosine methylation. The methyl group used for the methylation of both DNA and protein is derived from *S*-adenosyl methionine (SAM), which in turn derives its methyl group in a series of reactions from homocysteine utilizing methyltetrahydrofolate (MTHF) (Figure 7.2a). It is apparent from this biochemical pathway that an adequate supply of SAM is required for establishment of methylation marks and this could be influenced by nutritional factors as well as chemicals that interfere in the pathway.

One of the earliest and most influential studies in the nutrition field was that of the children of Dutch mothers malnourished during their first trimester in the postwar years of 1945–1946. Epidemiological studies of the children of these mothers have established that they have increased rates of obesity, and, in daughters, overall mortality, compared with those that might have been expected [52,53]. This is hypothesized to be due to an alteration in the epigenetic methylation marks on the DNA resulting from nutritional deprivation during the early stages of life when critical processes are underway to reestablish methylation patterns from the mother and father in the primordial germ cells. Indeed, further analysis of methylation imprinting in this cohort suggested the mechanism leading to increased obesity could be due to hypomethylation of the *Insulin Growth Factor II* (*IGF-2*) gene [54]. Demethylation of this gene could be detected six decades after the event suggesting that hypomethylation of *IGF2* can act as a past biomarker of early life malnutrition [54].

Among the most established compounds shown to cause epigenetic change are the metals, in particular arsenic. Arsenic exposure has occurred extensively

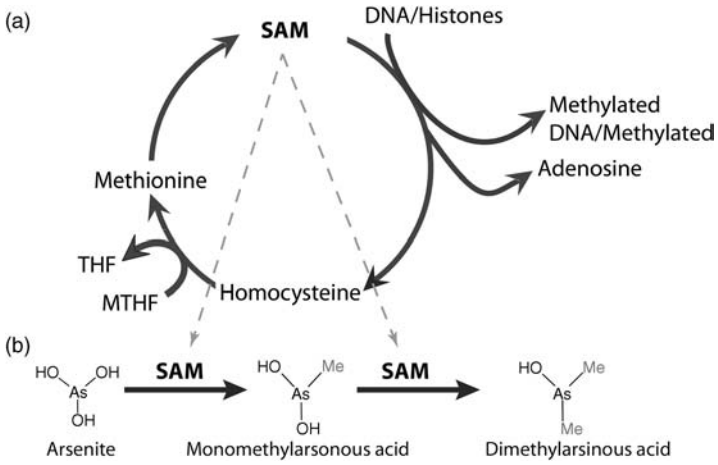


Figure 7.2 The SAM pathway (a) and metabolism of arsenite to monomethylarsonous acid and then dimethylarsinous acid using SAM (b). THF is tetrahydrofolate and MTHF is methyltetrahydrofolate.

in Southeast Asia, but also in other regions of the world as a result of water extraction from rock layers containing arsenic. It is estimated that more than 100 million individuals are exposed to levels greater than the 10 μg/dl permitted by the US Environmental Protection Agency (EPA) and recommended by the World Health Organization (WHO). As with nutritional effects, we have an understanding of the likely mechanism. Arsenic is metabolized first to a monomethyl form (monomethylarsonous acid) and then to a dimethyl form (dimethylarsinous acid) using SAM as a substrate (Figure 7.2b). It is therefore a small step to hypothesizing that arsenic exposure could lead to a depletion of SAM and therefore lack of substrate for the methylation of DNA. This has been demonstrated in both mice and rats where arsenic exposure led to a global hepatic DNA demethylation (hypomethylation) [55–58]. Similar effects have been seen in a number of cell lines reviewed by Ren *et al.* [59]. Conversely, a cohort study of 114 paired maternal and umbilical cord blood samples from mothers and infants in Southeast Asia showed an increased methylation at LINE-1 elements in both the maternal and cord leukocytes, which positively correlated with arsenic measured in the mother's urine [60]. This was hypothesized to be due to altered folate intake. A similar study in a Bangladeshi population that looked at histone modifications also showed ambiguous results, with increased histone methylation in females and decreased histone methylation in males. The authors hypothesized that this might have been due to the estrogen sensitivity of histone methylation marks [61]. Furthermore, the authors suggested that increased levels of the histone methylation mark H3K9me2 could lead to the increased methylation of LINE-1 elements that have been correlated with arsenic exposure.

Using the nongenotoxic carcinogen phenobarbital, Lempiäinen *et al.* demonstrated that increased hepatic expression of *Cyp2b10* mRNA is related to a specific hypomethylation in the promoter region of this gene [62]. Demethylation was confirmed to occur in the liver but not in the kidney, resulting in an increased expression of *Cyp2b10* mRNA in the liver but not the kidney. It is now necessary to build on this work and determine if other compounds in the same toxicological class as phenobarbital, namely, CAR agonists and nongenotoxic carcinogens, are able to similarly alter the promoter methylation and subsequent expression of *Cyp2b10*. If this promoter methylation is shown to have specificity for CAR agonists or nongenotoxic carcinogens, it could act as a specific and longer term epigenetic biomarker of exposure to these agents.

What is apparent is that methylation marks on DNA and histones have the potential to act as biomarkers of exposure, but there is some way to go before their potential is fully realized.

7.4
Novel Biomarkers: Beyond Nucleotide-Based Discovery

Much of the focus on biomarker discovery to date has centered on the identification of new protein- and nucleotide-based indicators of association, as detailed

here and elsewhere in this book. These biomarkers, together with other less well-explored targets and strategies, are likely to have a large impact on the development of more accurate, detailed, and informative dysfunction classification endpoints beyond the currently used diagnostic and toxicity assessment tools [63,64]. We will describe some of these new approaches covering a wide selection of strategies in order to highlight the diverse options that could be developed and integrated into future discovery programs.

Metabolites as biomarkers are not new. On the contrary, they were often the first discovered markers associated with disease and dysfunction. Serum levels of the metabolites creatinine [65] and bilirubin [66] are still used as clinical indicators of kidney and liver function, respectively. However, some of these metabolites may not be ideal for detecting subtle injury or disease, exemplified by the fact that considerable damage to the kidney needs to occur for changes in serum creatinine to appear. New approaches to identifying metabolites or metabolite profiles that better predict or associate with various levels of tissue injury or disease state are continuously being developed [67]. These involve ever increasingly sensitive platforms based around HPLC, MS, and NMR technologies and to date have identified important metabolic signatures of pathophysiology ranging from diabetes to environmental cadmium exposure [68–72]. Given that many injury biomarkers are released from damaged cells on cell death and represent an injury that has already occurred, the advantage of metabolomic screening may lie in the ability to detect metabolic profiles that represent changes in cellular function that occur prior to actual damage. Characterization of such profiles may allow better prediction of injurious conditions. Following on from this, technologies involving magnetic resonance imaging (MRI) and position emission tomography (PET) of radiolabeled metabolites to detect tissue injury and disease are a focus of many new approaches for biomarker evaluation. One such example involves the visualization of ^{13}C-labeled glutathione by MRI in human subjects and in rat models of liver injury to indicate oxidative stress within the tissue [73]. Imaging of phosphorous-containing metabolites in tissues using ^{31}P-MRS (magnetic resonance spectroscopy) has also been used to identify potential biomarkers of injury and disease [74,75].

Developments in technology are at the forefront of biomarker discovery, but their application to less well-characterized biological outputs may also be an interesting avenue for target discovery in the future. One example of such an approach is the increased analysis of exhaled air for biomarker characterization with exhaled nitric oxide and carbon monoxide among those currently being examined for their associative potential with exposure [76,77]. While these and many other well-studied analytes are continuing to be explored, it is clear that scope exists for the application of high-throughput technologies to screen for new biomarkers in exhaled air. This has been investigated for the analysis of metabolites/volatile organic compounds (VOCs) by GC/MS in the diagnosis of lung cancer with some success [78]. Indeed, electronic nose devices to detect these VOCs in the clinic are being developed, with the aim to provide fingerprints of different diseases by sampling exhaled air [79]. The application of other

technologies for similar analysis may bring further knowledge and better targets in the future. As previously discussed, extracellular or exosomal vesicles in biological fluids represent another potential source of biomarkers, which have not been fully explored to date. These vesicles in blood or urine, for example, can also contain specific protein molecules in addition to RNA that are indicative of specific injury or disease [80,81]. It is also important to consider older approaches to biomarker discovery that may not necessarily involve new technologies but involve new thinking surrounding pathophysiology and tissue injury. Such an approach has been used recently to suggest circulating giant macrophages as potential biomarkers of solid tumors [82] and circulating brain microvascular endothelial cells as biomarkers for blood–brain barrier injury [83]. Finally, we want to highlight the potential for integration of information from different high-throughput screening and characterization approaches as an important further step forward in biomarker discovery. Such approaches have been used to reveal mechanistic insights and more detailed understanding of biological responses that may allow for better selection of candidate biomarkers for future validation [84–86]. With modern analysis strategies, identification of changes in stress response pathways is a basic output from biomarker screening and integrative approaches [87,88]. Focusing on these changes represents an important goal for biomarker discovery in the future as these pathways have the potential to be used as predictors of tissue injury and disease before actual injury occurs.

7.5 Summary and Outlook

Transcriptomic biomarkers are extensively described in the literature, but few are validated. Although they have a role to play, they suffer from requiring a comparator for evaluation and a sample of tissue. Of arguably greater functionality will be circulating RNA species such as miRNAs. These have already shown their value as present biomarkers of effect and are likely to be of use in both clinical and toxicological studies. Structural changes in DNA have already proven their versatility as prognostic biomarkers for many disease states. The output of genome sequencing projects will provide a rich source of information that is likely to lead to many of these DNA changes being validated and adopted. Perhaps uniquely, differential DNA methylation has the potential to act as a biomarker of past exposure, a present biomarker of effect, and a future biomarker of a subsequent event or prognosis. Although the dynamic nature of these marks, cell specificity, and technical challenges of measurement mean that a substantial amount of research will likely be required before any DNA methylation changes are established as reliable biomarkers, their potential has already been demonstrated. Downstream of the genome, metabolites offer substantial opportunities for both single-molecule and complex biomarkers. The discovery and use of these is being driven forward, as with genetics, by technology development. Finally, there are the biomarkers beyond the micromolecular level that include

further macromolecules and cells. This chapter highlights the many exciting and varied options for the continued development of novel biomarkers that will help to improve the biomonitoring of exposure, effect, and prognosis within humans and human biomaterials.

References

1 Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

2 Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503–511.

3 Staudt, L.M. (2002) Gene expression profiling of lymphoid malignancies. *Annu. Rev. Med.*, **53**, 303–318.

4 Dyrskjot, L., Thykjaer, T., Kruhoffer, M., Jensen, J.L., Marcussen, N. *et al.* (2003) Identifying distinct classes of bladder carcinoma using microarrays. *Nat. Genet.*, **33**, 90–96.

5 van't Veer, L.J., Dai, H.Y., van de Vijver, M.J., He, Y.D.D., Hart, A.A.M. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530–536.

6 Glas, A.M., Floore, A., Delahaye, L.J., Witteveen, A.T., Pover, R.C. *et al.* (2006) Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics*, **7**, 278.

7 Pomeroy, S.L., Tamayo, P., Gaasenbeek, M., Sturla, L.M., Angelo, M. *et al.* (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, **415**, 436–442.

8 Zoon, C.K., Starker, E.Q., Wilson, A.M., Emmert-Buck, M.R., Libutti, S.K. *et al.* (2009) Current molecular diagnostics of breast cancer and the potential incorporation of microRNA. *Expert Rev. Mol. Diagn.*, **9**, 455–467.

9 Suter, L., Schroeder, S., Meyer, K., Gautier, J.C., Amberg, A. *et al.* (2011) EU Framework 6 Project: Predictive Toxicology (PredTox) – overview and outcome. *Toxicol. Appl. Pharmacol.*, **252**, 73–84.

10 Moggs, J.G., Tinwell, H., Spurway, T., Chang, H.S., Pate, I. *et al.* (2004) Phenotypic anchoring of gene expression changes during estrogen-induced uterine growth. *Environ. Health Perspect.*, **112**, 1589–1606.

11 Gant, T.W., Baus, P.R., Clothier, B., Riley, J., Davies, R. *et al.* (2003) Gene expression profiles associated with inflammation, fibrosis and cholestasis in mouse liver after griseofulvin. *Environ. Health Perspect.*, **111**, 847–853.

12 Tonge, D.P., Tugwood, J.D., Kelsall, J., and Gant, T.W. (2013) The role of microRNAs in the pathogenesis of MMPi-induced skin fibrodysplasia. *BMC Genomics*, **14**, 338.

13 Heinloth, A.N., Boorman, G.A., Foley, J.F., Flagler, N.D., and Paules, R.S. (2007) Gene expression analysis offers unique advantages to histopathology in liver biopsy evaluations. *Toxicol. Pathol.*, **35**, 276–283.

14 Bushel, P.R., Heinloth, A.N., Li, J., Huang, L., Chou, J.W. *et al.* (2007) Blood gene expression signatures predict exposure levels. *Proc. Natl. Acad. Sci. USA*, **104**, 18211–18216.

15 Lamb, J., Crawford, E., Peck, D., Modell, J., Blat, I. *et al.* (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**, 1929–1935.

16 Smalley, J.L., Gant, T.W., and Zhang, S. (2010) Application of connectivity mapping in predictive toxicology based on gene-expression similarity. *Toxicology*, **268**, 143–146.

17 Zhang, S.D. and Gant, T.W. (2008) A simple and robust method for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinformatics*, **9**, 258–268.

18 Caiment, F., Tsamou, M., Jennen, D., and Kleinjans, J. (2014) Assessing compound carcinogenicity *in vitro* using connectivity mapping. *Carcinogenesis*, **35**, 201–207.

19 Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.

20 He, L. and Hannon, G. (2004) microRNAs: small RNAs with a big role in gene regulation. *Nat. Genet.*, **5**, 522–531.

21 Hsu, P.W., Huang, H.D., Hsu, S.D., Lin, L. Z., Tsou, A.P. *et al.* (2006) miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res.*, **34**, D135–D139.

22 Pillai, R.S., Bhattacharyya, S.N., and Filipowicz, W. (2007) Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.*, **17**, 118–126.

23 Wang, K., Zhang, S.L., Marzolf, B., Troisch, P., Brightman, A. *et al.* (2009) Circulating microRNAs, potential biomarkers for drug-induced liver injury. *Proc. Natl. Acad. Sci. USA*, **106**, 4402–4407.

24 Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. *et al.* (2002) Identification of tissue-specific microRNAs from mouse. *Curr. Biol.*, **12**, 735–739.

25 Starkey Lewis, P.J., Dear, J., Platt, V., Simpson, K.J., Craig, D.G. *et al.* (2011) Circulating microRNAs as potential markers of human drug-induced liver injury. *Hepatology*, **54**, 1767–1776.

26 Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H. *et al.* (2010) The microRNA spectrum in 12 body fluids. *Clin. Chem.*, **56**, 1733–1741.

27 Callis, T. and Wang, D. (2008) Taking microRNAs to heart. *Trends Mol. Med.*, **14**, 254–260.

28 Fichtlscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A. *et al.* (2010) Circulating microRNAs in patients with coronary artery disease. *Circ. Res.*, **107**, 677–684.

29 Manolio, T.A. and Collins, F.S. (2009) The HapMap and genome-wide association studies in diagnosis and therapy. *Annu. Rev. Med.*, **60**, 443–456.

30 Anonymous (2010) Human genome: genomes by the thousand. *Nature*, **467**, 1026–1027.

31 Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C. *et al.* (2013) Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.*, **9**, e1003212.

32 Johansson, I. and Ingelman-Sundberg, M. (2011) Genetic polymorphism and toxicology – with emphasis on cytochrome P450. *Toxicol. Sci.*, **120**, 1–13.

33 Kerb, R., Hoffmeyer, S., and Brinkmann, U. (2001) ABC drug transporters: hereditary polymorphisms and pharmacological impact in MDR1, MRP1 and MRP2. *Pharmacogenomics*, **2**, 51–64.

34 Girirajan, S., Campbell, C.D., and Eichler, E.E. (2011) Human copy number variation and complex genetic disease. *Annu. Rev. Genet.*, **45**, 203–226.

35 Egan, C., Sridhar, S., Wigler, M., and Hall, I. (2007) Recurrent DNA copy number variation in the laboratory mouse. *Nat. Genet.*, **39**, 1384–1389.

36 Boysen, G., Pachkowski, B.F., Nakamura, J., and Swenberg, J.A. (2009) The formation and biological significance of N7-guanine adducts. *Mutat. Res.*, **678**, 76–94.

37 Loechler, E.L. (1995) How are potent bulky carcinogens able to induce such a diverse array of mutations? *Mol. Carcinog.*, **13**, 213–219.

38 BelguiseValladier, P. and Fuchs, R.P.P. (1995) *N*-2-Aminofluorene and *N*-2 acetylaminofluorene adducts: the local sequence context of an adduct and its chemical structure determine its replication properties. *J. Mol. Biol.*, **249**, 903–913.

39 Carthew, P., Rich, K.J., Martin, E.A., De Matteis, F., Lim, C.K. *et al.* (1995) DNA damage as assessed by ^{32}P-postlabelling in three rat strains exposed to dietary tamoxifen: the relationship between cell proliferation and liver tumour formation. *Carcinogenesis*, **16**, 1299–1304.

40 Hernandez-Ramon, E.E., Sandoval, N.A., John, K., Cline, J.M., Wood, C.E. *et al.*

(2014) Tamoxifen (TAM)–DNA adduct formation in monkey and human reproductive organs. *Carcinogenesis*, **35**, 1172–1176.

41 Ramesh, A. and Knuckles, M.E. (2006) Dose-dependent benzo(*a*)pyrene [B(*a*)P]–DNA adduct levels and persistence in F-344 rats following subchronic dietary exposure to B(*a*)P. *Cancer Lett.*, **240**, 268–278.

42 Bieler, C.A., Cornelius, M.G., Stiborova, M., Arlt, V.M., Wiessler, M. *et al.* (2007) Formation and persistence of DNA adducts formed by the carcinogenic air pollutant 3-nitrobenzanthrone in target and non-target organs after intratracheal instillation in rats. *Carcinogenesis*, **28**, 1117–1121.

43 Stephan-Blanchard, E., Chardon, K., Telliez, F., Arnould, J.P., Leke, A. *et al.* (2011) Are benzo[*a*]pyrene–DNA adducts an accurate biomarker of long-term *in utero* exposure to smoking? *Ther. Drug Monit.*, **33**, 329–335.

44 Goel, S., Evans-Johnson, J.A., Georgieva, N.I., and Boysen, G. (2013) Exposure profiling of reactive compounds in complex mixtures. *Toxicology*, **313**, 145–150.

45 Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, **447**, 425–432.

46 Zeybel, M., Hardy, T., Wong, Y.K., Mathers, J.C., Fox, C.R. *et al.* (2012) Multigenerational epigenetic adaptation of the hepatic wound-healing response. *Nat. Med.*, **18**, 1369–1377.

47 Szyf, M., Weaver, I., and Meaney, M. (2007) Maternal care, the epigenome and phenotypic differences in behavior. *Reprod. Toxicol.*, **24**, 9–19.

48 Youngson, N.A. and Whitelaw, E. (2008) Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.*, **9**, 233–257.

49 Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005) Epigenetic reprogramming in mammals. *Hum. Mol. Genet.*, **14** (Spec. No. 1), R47–R58.

50 Skinner, M.K., Manikkam, M., and Guerrero-Bosagna, C. (2010) Epigenetic transgenerational actions of environmental factors in disease etiology. *Trends Endocrinol. Metab.*, **21**, 214–222.

51 Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.

52 van Abeelen, A.F., Veenendaal, M.V., Painter, R.C., de Rooij, S.R., Dijkgraaf, M.G. *et al.* (2012) Survival effects of prenatal famine exposure. *Am. J. Clin. Nutr.*, **95**, 179–183.

53 Ravelli, G.P., Stein, Z.A., and Susser, M.W. (1976) Obesity in young men after famine exposure *in utero* and early infancy. *N. Engl. J. Med.*, **295**, 349–353.

54 Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J. *et al.* (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. USA*, **105**, 17046–17049.

55 Uthus, E.O. and Davis, C. (2005) Dietary arsenic affects dimethylhydrazine-induced aberrant crypt formation and hepatic global DNA methylation and DNA methyltransferase activity in rats. *Biol. Trace Elem. Res.*, **103**, 133–145.

56 Chen, H., Li, S., Liu, J., Diwan, B.A., Barrett, J.C. *et al.* (2004) Chronic inorganic arsenic exposure induces hepatic global and individual gene hypomethylation: implications for arsenic hepatocarcinogenesis. *Carcinogenesis*, **25**, 1779–1786.

57 Xie, Y., Trouba, K.J., Liu, J., Waalkes, M.P., and Germolec, D.R. (2004) Biokinetics and subchronic toxic effects of oral arsenite, arsenate, monomethylarsonic acid, and dimethylarsinic acid in v-Ha-ras transgenic (Tg.AC) mice. *Environ. Health Perspect.*, **112**, 1255–1263.

58 Okoji, R.S., Yu, R.C., Maronpot, R.R., and Froines, J.R. (2002) Sodium arsenite administration via drinking water increases genome-wide and Ha-ras DNA hypomethylation in methyl-deficient C57BL/6J mice. *Carcinogenesis*, **23**, 777–785.

59 Ren, X., McHale, C.M., Skibola, C.F., Smith, A.H., Smith, M.T. *et al.* (2011) An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. *Environ. Health Perspect.*, **119**, 11–19.

60 Kile, M.L., Baccarelli, A., Hoffman, E., Tarantini, L., Quamruzzaman, Q. *et al.* (2012) Prenatal arsenic exposure and DNA methylation in maternal and umbilical cord blood leukocytes. *Environ. Health Perspect.*, **120**, 1061–1066.

61 Chervona, Y., Hall, M.N., Arita, A., Wu, F., Sun, H. *et al.* (2012) Associations between arsenic exposure and global posttranslational histone modifications among adults in Bangladesh. *Cancer Epidemiol. Biomarkers Prev.*, **21**, 2252–2260.

62 Lempiäinen, H., Muller, A., Brasa, S., Teo, S.S., Roloff, T.C. *et al.* (2011) Phenobarbital mediates an epigenetic switch at the constitutive androstane receptor (CAR) target gene Cyp2b10 in the liver of B6C3F1 mice. *PLoS One*, **6**, e18216.

63 Boekelheide, K. and Schuppe-Koistinen, I. (2012) SOT/EUROTOX debate: biomarkers from blood and urine will replace traditional histopathological evaluation to determine adverse response. *Toxicol. Sci.*, **129**, 249–255.

64 Campion, S., Aubrecht, J., Boekelheide, K., Brewster, D.W., Vaidya, V.S. *et al.* (2013) The current status of biomarkers for predicting toxicity. *Expert Opin. Drug Metab. Toxicol.*, **9**, 1391–1408.

65 Shemesh, O., Golbetz, H., Kriss, J.P., and Myers, B.D. (1985) Limitations of creatinine as a filtration marker in glomerulopathic patients. *Kidney Int.*, **28**, 830–838.

66 Fevery, J., Van Damme, B., Michiels, R., De Groote, J., and Heirwegh, K.P. (1972) Bilirubin conjugates in bile of man and rat in the normal state and in liver disease. *J. Clin. Invest.*, **51**, 2482–2492.

67 Mamas, M., Dunn, W.B., Neyses, L., and Goodacre, R. (2011) The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch. Toxicol.*, **85**, 5–17.

68 Zhang, X., Choi, F.F., Zhou, Y., Leung, F.P., Tan, S. *et al.* (2012) Metabolite profiling of plasma and urine from rats with TNBS-induced acute colitis using UPLC–ESI-QTOF-MS-based metabonomics – a pilot study. *FEBS J.*, **279**, 2322–2338.

69 Park, Y.H., Lee, K., Soltow, Q.A., Strobel, F.H., Brigham, K.L. *et al.* (2012) High-performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. *Toxicology*, **295**, 47–55.

70 Hasegawa, M., Ide, M., Kuwamura, M., Yamate, J., and Takenaka, S. (2010) Metabolic fingerprinting in toxicological assessment using FT-ICR MS. *J. Toxicol. Pathol.*, **23**, 67–74.

71 Arora, P. (2014) Metabolomics yield a novel biomarker for predicting diabetes mellitus risk in humans. *Circ. Cardiovasc. Genet.*, **7**, 95–96.

72 Ellis, J.K., Athersuch, T.J., Thomas, L.D., Teichert, F., Perez-Trujillo, M. *et al.* (2012) Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.*, **10**, 61.

73 Skamarauskas, J.T., Oakley, F., Smith, F.E., Bawn, C., Dunn, M. *et al.* (2013) Non-invasive *in vivo* magnetic resonance measures of glutathione synthesis in human and rat liver as an oxidative stress biomarker. *Hepatology*. doi: 10.1002/hep.26925.

74 Solanky, B.S., Sanchez-Canon, G.J., Cobbold, J.F., Taylor-Robinson, S.D., Bell, J.D. *et al.* (2012) Metabolic profiling of the rat liver after chronic ingestion of alpha-naphthylisothiocyanate using *in vivo* and *ex vivo* magnetic resonance spectroscopy. *Toxicol. Sci.*, **126**, 306–316.

75 Fiorina, P., Bassi, R., Gremizzi, C., Vergani, A., Caldara, R. *et al.* (2012) ^{31}P-magnetic resonance spectroscopy (^{31}P-MRS) detects early changes in kidney high-energy phosphate metabolism during a 6-month Valsartan treatment in diabetic and non-diabetic kidney-transplanted patients. *Acta Diabetol.*, **49** (Suppl. 1), S133–S139.

76 Liu, F., Li, W., Pauluhn, J., Trubel, H., and Wang, C. (2013) Rat models of acute lung injury: exhaled nitric oxide as a sensitive, noninvasive real-time biomarker of prognosis and efficacy of intervention. *Toxicology*, **310**, 104–114.

77 Ryter, S.W. and Choi, A.M. (2013) Carbon monoxide in exhaled breath testing and therapeutics. *J. Breath Res.*, **7**, 017111.

78 Taivans, I., Bukovskis, M., Strazda, G., and Jurka, N. (2014) Breath testing as a method for detecting lung cancer. *Expert Rev. Anticancer Ther.*, **14** (2), 121–123.
79 Fens, N., van der Schee, M.P., Brinkman, P., and Sterk, P.J. (2013) Exhaled breath analysis by electronic nose in airways disease. Established issues and key questions. *Clin. Exp. Allergy*, **43**, 705–715.
80 Yang, X., Weng, Z., Mendrick, D.L., and Shi, Q. (2014) Circulating extracellular vesicles as a potential source of new biomarkers of drug-induced liver injury. *Toxicol. Lett.*, **225**, 401–406.
81 Zhou, H., Cheruvanky, A., Hu, X., Matsumoto, T., Hiramatsu, N. *et al.* (2008) Urinary exosomal transcription factors, a new class of biomarkers for renal disease. *Kidney Int.*, **74**, 613–621.
82 Adams, D.L., Martin, S.S., Alpaugh, R.K., Charpentier, M., Tsai, S. *et al.* (2014) Circulating giant macrophages as a potential biomarker of solid tumors. *Proc. Natl. Acad. Sci. USA*, **111**, 3514–3519.
83 Huang, S.H., Wang, L., Chi, F., Wu, C.H., Cao, H. *et al.* (2013) Circulating brain microvascular endothelial cells (cBMECs) as potential biomarkers of the blood–brain barrier disorders caused by microbial and non-microbial factors. *PLoS One*, **8**, e62164.
84 Matheis, K.A., Com, E., Gautier, J.C., Guerreiro, N., Brandenburg, A. *et al.* (2011) Cross-study and cross-omics comparisons of three nephrotoxic compounds reveal mechanistic insights and new candidate biomarkers. *Toxicol. Appl. Pharmacol.*, **252**, 112–122.
85 Forgacs, A.L., Kent, M.N., Makley, M.K., Mets, B., DelRaso, N. *et al.* (2012) Comparative metabolomic and genomic analyses of TCDD-elicited metabolic disruption in mouse and rat liver. *Toxicol. Sci.*, **125**, 41–55.
86 Wilmes, A., Limonciel, A., Aschauer, L., Moenks, K., Bielow, C. *et al.* (2013) Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress. *J. Proteomics*, **79**, 180–194.
87 Jennings, P., Limonciel, A., Felice, L., and Leonard, M.O. (2013) An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.*, **87**, 49–72.
88 Jennings, P. (2013) Stress response pathways, toxicity pathways and adverse outcome pathways. *Arch. Toxicol.*, **87**, 13–14.

8
Predictive Toxicology: Genetics, Genomics, Epigenetics, and Next-Generation Sequencing in Toxicology

Tobias Heckel and Laura Suter-Dick

8.1
Introduction

Genetics is the study of the genome with information stored in the DNA sequence of an organism. Under normal conditions, this information is considered to remain stable throughout the life span of an individual. However, there are multiple outcomes coded by these sequences that enable the development of a complex organism from a single totipotent cell through a series of exquisite spatiotemporal regulation steps. Thus, for a thorough understanding of molecular mechanisms in biological processes including toxicological studies, there is a need to understand changes in the genetic information as well as in its regulation. This requires the study of the DNA sequence as well as of its regulatory mechanisms, taking into consideration interindividual and interspecies differences. The availability of reference genomes allows systematic comparisons across species for any gene or protein involved in pharmacology or toxicology. They also provide the basis for the development and application of tools for genome-wide expression profiling and association studies of single-nucleotide polymorphisms (SNPs), deletions, amplifications, or epigenetic modifications. Several genes and pathways are involved in susceptibility to adverse events. In particular, the study of intraspecies differences in metabolism is at the forefront of genotyping studies, since they are often directly associated with side effects in humans.

Many of the genetic control mechanisms are achieved through epigenetic changes, a concept dating from 1942 to describe the differentiation of cells with the same genetic information from their initial totipotent state to highly specialized cell types during the embryonic development. It is now well recognized that the regulation of the gene expression by epigenetic mechanisms is influenced not only during the embryological development but also under pathological conditions and as a result of toxic insult. Molecular mechanisms leading to these epigenetic effects include DNA methylation, chromatin remodeling, and regulation through noncoding RNAs (ncRNAs, such as microRNAs (miRNAs)). In toxicology, there has been an increasing interest in the interactions of xenobiotics with the genome (e.g., specific polymorphisms that influence susceptibility and

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

metabolism), the transcriptome (e.g., toxicogenomic evaluations), and the epigenome (e.g., DNA methylation and chromatin conformation changes).

The advances in understanding the genetic and epigenetic phenomena under normal and toxicological/pathological conditions and their consequences in an organism have recently been made possible by the technological advances in molecular biology. The new technologies and the increased understanding of the molecular mechanisms underlying toxicological events have enabled the fast evolution of molecular toxicology. After the advent of the microarray in the 1990s, which made toxicogenomic research possible, next-generation sequencing (NGS) enables now the sequencing of a genome or a transcriptome.

8.2 Technological Advances

Moore's law is the computer industry's trend of doubling the number of transistors on an integrated circuit and thus the computing power every 2 years. Until the year 2008, continuous advances in DNA sequencing technology reduced the sequencing costs by a factor of 2 or 3 each year, only to outpace Moore's law thereafter [1]. Somewhat analogous to miniaturization and integration of electronic components in microprocessors, miniaturization, parallelization, and integration enabled DNA sequencing systems to work at greater speed, at lower cost, and with more samples and less reagents.

Until the introduction of NGS systems, Sanger's chain termination chemistry was the most widely used sequencing method. It improved gradually over three decades introducing fluorescently labeled nucleotides, parallelization of electrophoresis in independent capillaries, and general automation [2]. Prerequisite for Sanger sequencing is an amplified DNA template from plasmids or PCR amplicons. Iterative cycles of template denaturation, primer annealing, primer extension, and stochastically incorporation of fluorescently labeled, chain-terminating dideoxynucleotides (ddNTPs) generate a mixture of end-labeled extension products, which are size separated in a capillary-based polymer gel. The resulting sequence ladder has an average length of 700 bases. Raw accuracies of Sanger sequencing are between 99.4% [3] and 99.999% per base [4,5]. This performance is the reason why this method is still frequently used for small-scale validation studies, despite its low throughput.

For higher throughput studies, Sanger sequencing has been replaced by next-generation sequencing strategies that were built based on miniaturization in microfluidic flow cells or picoliter chambers allowing parallelization of millions of sequencing reactions. The DNA template for these sequencing reactions is sheared into fragments and then clonally amplified by emulsion PCR on beads (Roche/454, Ion Torrent, ABI/SOLiD) or bridge PCR on surface-linked oligonucleotides (Illumina). Sequencing reactions are spatially separated through immobilization in picotiter or microwell plates (Roche/454, Ion Torrent) or onto flow cell surfaces (Illumina, ABI/SOLiD).

The systems of Roche/454 [3] and Ion Torrent [6] use the sequencing by synthesis method generating either a chemiluminescence signal or a hydrogen ion-sensitive potential change. This signal is proportional to the number of incorporated nucleotides, which in turn makes these systems prone to homopolymer-length sequencing errors (insertion–deletion errors) [7,8]. The per-base raw accuracies are around 99.5% for Roche/454 [8,9] and 98.897% for Ion Torrent within the first 100 bases [6].

The Illumina sequencing system [10,11] is based on single-base extensions with fluorescently labeled nucleotides that are also reversible chain terminators. In this case, after each sequencing cycle, the fluorophore is removed, and the nucleotide can be extended in ensuing cycles. The limitations of this system are shorter read lengths and a lower raw accuracy of 98.5% [5] concomitant with a higher susceptibility to single-base substitution errors. These errors can be caused by incomplete extension and insufficient removal of reversible terminators [12–15].

The ABI SOLiD system uses DNA ligase to sequentially incorporate several fluorescently labeled oligonucleotide octamers [5,16]. Progressive rounds of octamer ligation enable sequencing of two adjacent bases central to the octamer rather than one. In addition, each base is interrogated in two different ligation reactions. Although substitutions are the most common error type of the SOLiD system, its two-base encoding method contributes to high raw accuracies of 99.94% [17] at the limitation of very short read lengths.

The next technology leap in next-generation sequencing is single-molecule sequencing using DNA templates without any amplification. Pacific BioSciences has commercialized a single-molecule real-time sequencing (SMRT) by synthesis technology recording the incorporation time of fluorescently labeled nucleotides into single surface immobilized DNA polymerases [18]. These polymerases are Poisson-like distributed over a SMRT cell, a chip consisting of thousands of nanometer-scale wells. Benefits of the SMRT system are long read lengths of thousands of bases in one stretch, the high sequencing speed, and in principle the ability to detect chemical DNA modifications such as methylation [19]. The per-base raw accuracy of this method is low with 83% due to extremely short interphase intervals between nucleotide incorporation events [20]. Low yield ratios of polymerase-occupied wells further limit the throughput.

All next-generation sequencing platforms benefited tremendously by miniaturization and massive parallelization, whereas the different systems trade off the length of sequenced nucleotide reads for the total number of reads that can be simultaneously acquired. These simultaneous measurements are key features of NGS technologies, because they allow not only qualitative sequence analysis but also digital counting of the sequenced reads. In general, the rapid progress in NGS technologies made this process so fast, robust, and cost effective that it increases the feasibility of a wide spectrum of systematic biological studies analyzing genome and genetic variation, RNA expression, splice variant expression, translation of mRNA sequences bound to ribosomes (ribosome profiling), protein–DNA interactions, epigenetic DNA modifications, and DNA/chromosomal structural variations.

Since NGS methods are confined to the analysis of DNA sequences, RNA or epigenetic sequencing applications require more sophisticated pretreatments of samples before sequencing. For RNA sequencing, several methods exist that enrich or deplete certain RNA species before reverse transcription into DNA and subsequent sequencing. Those methods employ, for example, polyA-based capture beads for mRNA enrichment, selective biotinylated probes and streptavidin beads to remove ribosomal RNA, or RNase incubation and size exclusion chromatography to purify ribosome-protected RNA fragments [21]. For epigenetic applications, affinity-based enrichment techniques and bisulfite treatments are most commonly used. Affinity enrichment techniques employ methyl-binding domain (MBD)-containing proteins, chemical tagging, for example, streptavidin purification of selectively labeled hydroxymethylcytosines, or antibodies specific for epigenetic modifications, such as 5-methylcytosine or 5-hydroxymethylcytosine. Although similar in approach, these enrichment methods target different regions of the genome. While immunoprecipitation capture approaches are biased to bind more methylated regions with low density of cytosine–guanine dinucleotide (CpG)-rich sequences, MBD-based approaches favor high-CpG-density regions. However, inclusion of multiple elution steps with increasing salt concentrations in MBD capture protocols can enrich moderately methylated regions, making it more useful compared with immunoprecipitation [22]. Bisulfite treatment or oxidative bisulfite treatment is the primary technology used to identify 5-methylcytosine or 5-hydroxymethylcytosine, respectively, because of its single-base or regional resolution in the genome. By using optimal conditions, unmethylated cytosines are deaminated to uracils, while methylated cytosines remain unconverted. Subsequent PCR amplification will replace all uracils derived from unmethylated cytosines with thymine, and in this way an epigenetic methyl mark is converted into a genetic difference detectable by next-generation sequencing [22,23]. However, this method is costly and effort intensive, especially for large number of samples.

8.3 Applications in Toxicology

8.3.1 Genome Sequencing and Sequence Level Comparisons

The fact that all living organisms use the same oligonucleotide chemistry, be it DNA or RNA, makes genomics a versatile tool that can be applied to any individual or species relevant for toxicology. This is the reason why genomics is perceived as having the greatest impact on toxicology in the identification of species and interindividual differences.

The biggest success of genomics, and especially of next-generation sequencing technologies, is in the field of genome sequencing since cost and time required to sequence one haploid genome (~3 Gb) have dropped dramatically after

Table 8.1 Overview on genome sequencing projects of humans and different animal species.

Technology	Project/Species	Year(s)	Cost	References
Sanger sequencing	Human Genome Project	1990–2003	USD 3×10^9	[24]
	Mouse (C57BL/6J)	2002	USD 130.5×10^6	[25]
	Rat (Brown Norway)	2002	USD 118.5×10^6	[26]
	Boxer dog	2005	USD 30×10^6	[27]
	Rhesus monkey	2007	USD 22×10^6	[28]
	Common marmoset	2010	Undisclosed	[29]
Next-generation sequencing	Human (James Watson)	2008	USD 1.5×10^6	[30,31]
	Cynomolgus monkey	2011	USD <100 000	[20,32,33]
	Beagle dog	2013	Undisclosed	[34]
	Ellegaard Göttingen minipig	2013	Undisclosed	[34]

completion of the Human Genome Project (Table 8.1). Thus, next-generation sequencing paved the way to generate comprehensive sets of reference genomes for humans and all major animal models used in preclinical research to identify differences at the level of genes, predicted proteins, and genome architecture (Figure 8.1).

Comparing orthologous sequences of DNA, mRNA, and proteins across multiple species at varying evolutionary distances allows the identification of sequences that are unique for a given species as well as of those that are conserved between experimental models and humans. Such ortholog comparisons have proven useful in drug target prediction mainly in the fields of pharmacology but also in toxicology, as they enable the study of translational aspects. They are based on the assumption that evolutionary conservation of sequences correlates with conservation of function [35]. Since sequence similarities provide only limited inferences about gene and protein functions across species, additional knowledge about functional domains, protein structure, gene regulatory sites, and binding affinities is needed.

Furthermore, tissue- and cell type-specific expression levels in the adult as well as during the development should be integrated whenever possible to select the most relevant preclinical models. This integrated knowledge can help to discriminate better between responder and nonresponder species, and to uncover potential mechanisms of exaggerated pharmacology, lack of efficacy, and toxicity. For example, comparative genetics has been performed on genes coding for possible pharmacological targets. One of them is the trace amine-associated receptor 1 (TAAR1) that has proven to be an important modulator of the dopaminergic system and is considered a promising target for the treatment of

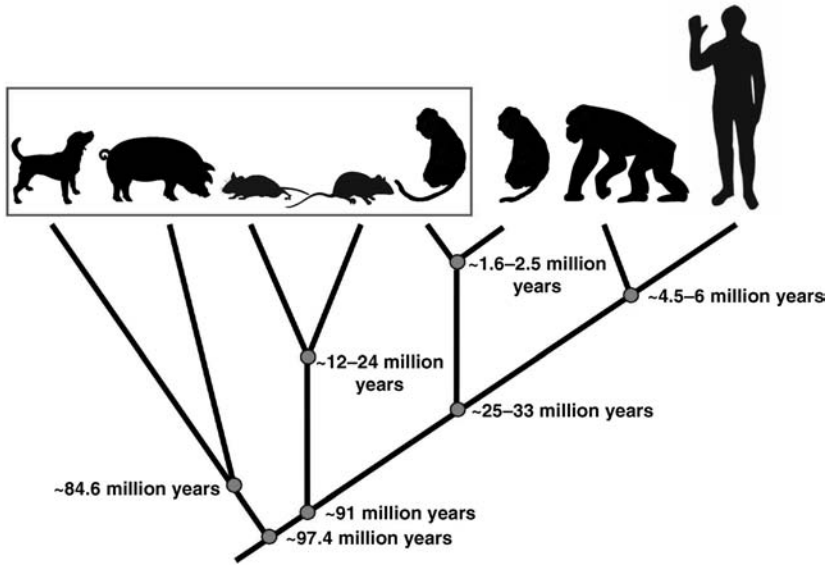


Figure 8.1 Reference genomes and estimates of evolutionary divergence times. Comprehensive sets of reference genomes for humans and all major preclinical animal models (box) allow the assessment of species similarities at the level of genes, predicted proteins, gene regulatory elements, and genome architecture. Animals from left to right: beagle dog, minipig, mouse, rat, cynomolgus monkey, rhesus monkey, and chimpanzee. Estimates of evolutionary divergence time based on sequence identity are indicated at each node. Values taken from Refs [25,26,28,32] and www.timetree.org/.

neuropsychiatric disorders. Sequencing of the TAAR1 gene revealed incomplete and poorly conserved coding regions between the dogs and humans, due to several frameshift mutations resulting in an uncompensated loss of function. This finding has a profound impact in the selection of species for pharmacological and toxicological investigations, since it disqualifies the dog as a pharmacological model [36].

Another type of comparative study compared the DNA sequences of dioxin response elements (DREs) – binding sites for the aryl hydrocarbon receptor (AhR) – in human, mouse, and rat genomes. Such gene regulatory elements tend to be conserved through evolution for common expression responses, but the integrated analysis of positionally conserved DRE sequences showed interspecies differences between rodents and humans: of the mouse–rat orthologous genes with a DRE close to transcription start sites (+/− 1500 bp), only 37% had an equivalent human ortholog. These results suggest that AhR-mediated gene expression may differ between species, which has direct implications in human risk assessment [37]. These results enable the prediction of species-specific differences in responses and enable toxicologists to evaluate the data in a relevant context to perform scientifically based human risk assessment.

Furthermore, sequence comparisons across individuals in specific target genes and in genome-wide association studies (GWAS) have been very successful in

identifying interindividual genetic differences that predict disease susceptibility or drug response not only in human populations, but also in animals. In humans, GWA studies on drug-induced liver injury (DILI) identified human leukocyte antigen (HLA) variants as risk factors for hepatotoxicity with the antibiotic flucloxacillin. These variants resulted from data obtained from 51 DILI cases and 282 controls for which around 1 million SNPs were genotyped [38]. It is also known in dogs that certain breeds are prone to develop spontaneously latent diseases. This fact is important for toxicologists to discriminate findings of spontaneous diseases from drug-induced injuries, especially in cases where a genetic predisposition may influence the outcome of a study.

In this context, recent GWA studies have been very successful in dogs due to the fact that the majority of dog breeds have been created few hundred years ago. Such a breeding system leads to large-size haplotype blocks within a breed due to extensive linkage disequilibrium. This in turn makes dog GWAS easier to perform, as it requires smaller sample size and less marker SNPs than human studies [39]. A recent study with 81 affected dogs and 57 controls and only around 15 000 SNPs identified five loci associated with canine systemic lupus erythematosus and steroid responsive meningitis–arteritis (SRMA) in the Nova Scotia duck tolling retriever [40]. Since SRMA, which is also known as beagle pain syndrome, is overrepresented not only in the Nova Scotia duck tolling retriever but also in other dog breeds including the pharmacologically relevant beagles, such loci can serve as predictive genetic biomarkers so that dog cohorts can be tested and stratified regarding disease predisposition.

Finally, genetic background is a more and more important experimental variable to explain interindividual variability in cynomolgus monkeys, one of the most widely used nonhuman primate species in biomedical research and drug testing. These animals live in highly diversified and genetically distinct continental and insular areas where they are captured and used to found stocks in breeding centers [41,42]. Their geographic origin has been found to be associated with phenotypes such as variability in tissue expression profiles [32], different immune genetics [43], susceptibility to retroviral infection [44], and spontaneous origin-dependent pathologies [45–47]. Thus, thorough understanding of findings in the cynomolgus monkey must take into consideration the genetic background of the animals. An example of the influence of genetic background has been published by Liu *et al.*; their data showed that polymorphisms of the CD3epsilon T-cell coreceptor in cynomolgus monkeys result in variable binding affinities of CD3epsilon targeting biologics. This had a major impact in the interpretation of the data as it confounded the determination of the lowest observed effect level (LOEL) [48].

8.3.2
Genotype and Metabolism

Interindividual differences in expression or function of genes responsible for drug metabolism and transport are caused by genetic, epigenetic, physiological,

and environmental factors. Primarily screening of target genes for polymorphisms, but also increasingly genome-wide association studies have improved our understanding of the contribution of genetic polymorphisms to the high variability in metabolism.

Polymorphisms in drug metabolizing enzymes and transporters can have considerable impact on the pharmacokinetic properties of drugs and have been best studied in humans to optimize drug treatment, particularly in the areas of oncology, cardiovascular disease, infection, and psychiatry [49]. These polymorphisms affect protein expression, activity, or both as illustrated in the following examples. Many drugs are metabolized by cytochrome P450 (CYP) enzymes. Within this large protein family, polymorphisms have been identified for CYP2B6, CYP2C9, CYP2C19, and CYP2D6 [50,51].

For example, a set of SNPs for the CYP2B6 gene, the CYP2B6*6 haplotype, leads to an enzyme with slower metabolic activity. As a consequence, CYP2B6*6 homozygous HIV patients treated with the antiretroviral reverse transcriptase inhibitor efavirenz had high plasma concentrations, which are associated with central nervous system (CNS)-related adverse reactions. In this case, a genotype-based dose reduction would be beneficial for the patients and greatly improve the clinical outcome [52].

The antiplatelet agent clopidogrel is a prodrug mainly activated by CYP2C19. Carriers of the CYP2C19*17 haplotype are found to have increased enzyme activity concomitant with enhanced response to treatment after acute myocardial infarction, but at the cost of increased risk of bleeding. Contrarily, carriers of defective CYP2C19 alleles, such as the CYP2C19*2 haplotype, show poor prodrug metabolism, resulting in a black-box warning by the US Food and Drug Administration (FDA) with respect to reduced effectiveness in homozygous patients [53].

A similar trade-off between efficacy and safety is well described for the vitamin K1 epoxide reductase (VKORC1) inhibiting anticoagulant warfarin. Here candidate gene and GWA studies have consistently shown that about 30–40% of the total interindividual variation in the final warfarin dose can be explained by genotypes of the drug target VKORC1 and the main metabolizing enzyme CYP2C9 [54].

After an FDA recommendation on genotype-specific dose ranges, the effectiveness of this predictive dosing strategy is currently being investigated in clinical trials [55].

In addition to anticoagulants, antidepressants and antipsychotics are prominent examples of drug classes for which genetic polymorphisms significantly affect pharmacokinetic parameters, which are partly associated with the 30–40% failure rate with initial depression treatment [56]. Many antipsychotics and antidepressants are known CYP2D6 substrates and plasma levels of these drugs at the same dosage can vary 5–20-fold among individuals. This variability commonly results from splice site defects, small deletions or nonsynonymous polymorphisms associated with the phenomenon of poor or intermediate metabolizers, and, in rarer cases, from CYP2D6 copy number variations of ultrarapid metabolizers. With respect to tricyclic antidepressants such as amitriptyline and nortriptyline, which are mixed serotonin and norepinephrine reuptake

inhibitors, CYP2D6 ultrarapid metabolizers have a higher probability of therapy failure due to subtherapeutic plasma concentrations.

Differently, CYP2D6 poor metabolizers are more likely to suffer from adverse effects such as anticholinergic, CNS, or cardiac effects due to elevated tricyclic plasma concentrations. Since CYP2D6 genotypes are highly predictive for enzyme activity and success of therapy, guidelines for dose adjustment or recommendations of alternative agents on the basis of the genotype have been formulated, but have not yet been tested prospectively in clinical trials [56]. Besides prominent CYP polymorphisms, clinically relevant genetic variants have also been found in other drug metabolizing enzymes and in transporters.

Variants in the promoter region of the drug metabolizing enzyme UGT1A1 result in a dose adjustment recommendation by the FDA for the DNA topoisomerase I inhibitor irinotecan used for the treatment of colon cancer. Individuals with the common UGT1A1*28 haplotype express lower enzyme levels and therefore have a reduced capability to detoxify the active metabolite of this prodrug. Since many anticancer drugs are used near cytotoxic doses, these drugs usually have a narrow therapeutic range so that patients are at higher risk of suffering an adverse drug reaction such as severe diarrhea and neutropenia [57]. Interestingly, the UGT1A1*28 haplotype was also shown to be associated with elevated unconjugated bilirubin levels in patients receiving tocilizumab, a monoclonal antibody inhibiting the IL-6 receptor for the treatment of rheumatoid arthritis. Due to this genetic association, the observed hyperbilirubinemia is with high probability not related to drug-induced hepatotoxicity, but to underlying pharmacodynamic effects. These effects probably arise from tocilizumab-mediated inflammation suppression, which in turn can impact UGT1A1 expression levels, especially in patients with the UGT1A1 variant, hampering the UGT1A1-mediated elimination of bilirubin. Another hypothesis is that increased hemoglobin production in tocilizumab-treated patients with the variant may lead to increased bilirubin [58].

In the case of the cholesterol-lowering drug simvastatin, candidate gene and retrospective GWA studies have revealed a strong association between a single SNP in the organic anion transporter gene (SLCO1B1) and drug-induced myopathy. This finding led to the creation and confirmation of a mechanistic link to additional SLCO1B1 amino acid changing SNPs, which cause impaired statin uptake into hepatocytes, reduced hepatobiliary excretion, higher plasma levels of statins, and myopathy due to statin-mediated interference with the mevalonate pathway in muscle [59,60].

Laboratory animal species also exhibit polymorphisms affecting several drug metabolic pathways, but this has been much less extensively researched than in humans. In rats, interstrain variability in drug metabolism has been observed in Dark Agouti, Lewis, and Sprague Dawley strains proposing them as models for human CYP2D6 poor (Dark Agouti) and extensive (Lewis, Sprague Dawley) metabolizer phenotypes, respectively [61,62]. Furthermore, interindividual variability in the Wistar strain has been associated with a CYP2D3 genetic polymorphism and differences in the metabolism of diazepam, a positive allosteric modulator of $GABA_A$ receptors [62].

In beagle dogs, genetic polymorphisms have been identified in a group of drug metabolizing CYPs, of which CYP1A2 polymorphisms have been shown most impactful on pharmacokinetics. In this case, large interindividual differences in the plasma concentrations of the phosphodiesterase type IV inhibitor, YM-64227, could be attributed to a beagle CYP1A2 deficiency genotype [63].

In cynomolgus monkeys, Jacqz *et al.* have reported a polymorphic phenotype with a frequency of poor metabolizers of 14% for the antihypertensive drug debrisoquine, probably mediated by polymorphic cynomolgus CYP2D17 and CYP2D44 isoforms, which are similar to human CYP2D6 [64,65].

Due to the advancement in next-generation sequencing technologies, the comprehensive characterization of animal genotype–phenotype associations is about to change as shown by a recent comparison of genomes of 28 rat strains. This study led to the identification of disease loci that overlap with previously mapped loci for related traits in humans, indicating shared pathways underlying these phenotypes in rats and humans [66]. Thus, the study of interspecies, interindividual, or interstrain differences as a surrogate for interindividual differences can be of value to bridge the gap between animal research and human risk assessment and to determine human relevance not only for disease models, but also for metabolism of drugs and toxic substances.

8.3.3 Mechanistic Toxicology and Toxicogenomics

As illustrated in the preceding section, sequence and expression changes in known key genes coding for proteins involved in xenobiotic metabolism can have a major impact on the toxicity outcome. Thus, these genes have been thoroughly investigated and susceptibility genotypes have been identified to treat patient populations adequately. However, the reductionist approach focusing on well-described, distinct phenotypes with high penetrance is not suited for the investigation of unexpected toxicity. Unpredicted toxicities usually involve complex biological processes in which several organs and pathways are affected in a yet not fully elucidated manner. Many toxicological findings fall into this category and can therefore not be sufficiently addressed by the study of single genes. Toxicogenomics, defined as the parallel measurement of a high number of transcripts, is therefore a suitable approach for the comprehensive analysis of toxicity-related drug-induced changes in a given tissue.

The widespread use of microarray technologies in the 1990s and 2000s delivered large data sets of gene expression changes associated with toxic insult, some of which were part of large undertakings in industry, regulatory bodies, and consortia such as the European Predictive Toxicology Project (PredTox) or Japanese Toxicogenomics Project (TGP) with its database TG-GATES (Toxicogenomics Project – Genomics Assisted Toxicity Evaluation System). The data sets allowed the generation of predictive models using advanced multivariate analysis and biostatistical tools (for a complete listing, see Chapter 6). However, the causal relationships between the toxicity phenotype and the modulation of

gene expression could not be elucidated as readily as expected because the functional relationship of many of these transcripts was not known.

Further developments in the field of toxicogenomics led to the incorporation of additional technologies such as proteomics and metabolomics as well as bioinformatic network modeling approaches, to get a better grasp on the functional and biological consequences of drug-induced transcriptional changes. The integrated analysis of these endpoints led to the generation of data-based mechanistic hypothesis involving a discrete number of relevant pathways or major hubs in networks. These identified key players in pathways or networks can then be studied in a more targeted approach and therefore provide biologically meaningful mechanisms.

Specific examples of the use of genomic technologies for the rational selection of pathways and mechanistic toxicology have been published, several of them addressing liver hypertrophy. This phenotype is relatively common in response to xenobiotic exposure but remains of relevance as it is often associated with liver tumor formation in the 2-year carcinogenicity studies in rodents. In the work published by the PredTox Consortium, for example, gene expression analysis identified two distinct mechanisms underlying the common liver hypertrophy phenotype:

1) A marked increase in the expression of xenobiotic metabolizing enzymes (XMEs) leading to proliferation of smooth endoplasmic reticulum (SER).
2) A marked upregulation of genes involved in peroxisomal fatty acid oxidation, associated with peroxisome proliferation.

Although both mechanisms were not mutually exclusive, liver hypertrophy was caused mainly by the proliferation of either SER or peroxisomes [67]. Similarly, transcriptomic data led to the identification of Cyp2b10 mRNA increase as biomarker for rodent-specific constitutive androstane receptor (CAR) activation accompanied by liver weight increases and hepatic tumors as observed, for example, after long-term exposure to the cholesterylester transfer protein (CETP) inhibitor dalcetrapib [68]. The results support the use of genomic technologies to foresee the outcome of such long-term studies, as the induction of Cyp2b10 can be observed early – in general after 2-week *in vivo* exposure – and is fairly predictive of the long-term outcome in the liver. In addition to the identification of relevant pathways and biomarkers in the liver, toxicogenomic platforms are useful to identify pathways associated with organ toxicity. This is particularly interesting when searching for specific biomarkers or translating *in vivo* and *in vitro* results within or across species. In an interesting study characterizing two histamine H3 receptor inverse agonists, the transcriptomic data generated after *in vivo* exposure of rats identified a strong effect on the cholesterol synthesis pathway accompanied by the prediction of liver toxicity. A subset of relevant genes was thus analyzed using qRT-PCR in the livers of the exposed animals as well as in rat and human hepatic cell cultures. The dysregulation of cholesterol synthesis, drug metabolism, and glutathione metabolism in rats

in vivo could be confirmed. Moreover, key enzymes of the cholesterol synthesis pathway, such as squalene epoxidase and HMG-CoA reductase (HMGCR), were induced in the cultured hepatocytes of both humans and rats. In summary, the results showed good concordance between *in vivo* and *in vitro*, were used for safety assessment, and thus led to the discontinuation of the compound, as the findings in the rodents are likely to be relevant to humans [69].

Thus, 20 years after the boom of toxicogenomics, researches have generally moved away from its use as a black-box predictive tool to a more scientifically driven tool to identify key biomarkers, pathways, or major hubs of networks that can be understood, bear biological relevance, and can be used as indicators of a toxicological response across biological systems. This change in mindset, together with technological advances from microarray to next-generation sequencing technologies (described above), has also promoted the incorporation of additional molecular parameters that aid the identification and understanding of toxicologically relevant events, for example, the assessment of the expression of miRNAs, DNA methylation, or other epigenetic changes.

8.3.4 Epigenetic Changes and miRNAs

There is an increasing body of evidence for the involvement of epigenetic perturbations in human disease and in response to environmental factors such as diet and exposure to xenobiotics. This fact adds a dimension of complexity to the study of adverse events, since persistent changes in DNA methylation and histone modification patterns may elicit a disease phenotype or an adverse event that becomes evident later in time. Thus, the study of acute and subchronic toxicity may not be suitable to reflect all possible adverse outcomes.

DNA methylation changes have been associated with several diseases such as schizophrenia, lupus erythematosus, type 2 diabetes, and cardiovascular diseases besides the well-established role of aberrant DNA methylation in cancer. The fact that the DNA methylation status is a dynamic process affecting postmitotic cells in the adult organism as well as proliferating and differentiating cells during embryogenesis has greatly expanded the possible implications in toxicology [70]. This fact is promoting the development of new approaches for the assessment of the DNA methylation status, such as genome-wide or locus-specific molecular methylation analysis methods or 3D quantitative DNA methylation imaging, which combines the detection of demethylated DNA with the phenotypic consequence of chromatin decondensation [71].

Although at present little data are available, some substances such as ethanol, inorganic arsenic, the DNA methylation inhibitor 5-azacytidine, or the anticonvulsant drug and histone deacetylase inhibitor valproic acid are known to cause DNA hypomethylation. Also, cigarette smoke has been shown to induce the demethylation of the oncogene synuclein-gamma through the downregulation of the DNA methyltransferase DNMT3B in lung cell lines, leading to a metastatic phenotype [72]. Of great concern is the fact that DNA methylation changes

might have a long-lasting or even irreversible impact on an organism, including cellular, physiological, and behavioral effects. This is highly relevant for the safety assessment of pharmaceuticals specifically interfering with DNA methylation and histone deacetylation, which could cause unspecific silencing of tumor suppressor genes, dysregulated oncogene expression, or submicroscopic genomic rearrangements via hypomethylated repetitive elements. Hence, epigenetics becomes particularly important in the study of cancer, since the initiating events leading to carcinogenesis may include temporal accumulation of genetic and epigenetic changes. Classic mechanisms inducing tumors involve direct interaction with DNA and cause changes in the DNA sequence (mutagens) or in the DNA structure (clastogens). Other substances might interfere with the chromosomal segregation during cell division and lead to an uneven distribution of the chromosomes in the daughter cells (e.g., aneugens).

All these mechanisms of genotoxicity are reasonably well understood as direct damage of the DNA sequence or structure is causative of the event. In addition to the agents causing direct DNA damage, nongenotoxic carcinogens are much less understood. Here, the initiation effects involve transcriptional regulation as well as epigenetic mechanisms such as DNA methylation changes, histone post-translational modifications, and altered gene regulation by ncRNAs.

Thus, nongenotoxic carcinogens are much more difficult to detect and their molecular mechanisms are still not fully understood. One of the most studied examples is the anticonvulsant drug and partial $GABA_A$ receptor agonist phenobarbital (PB), a very well-known nongenotoxic carcinogen that causes liver tumors in rodents after long-term exposure (e.g., in mandatory rodent carcinogenicity studies) [73]. The understanding of the underlying molecular mechanisms of these types of substances is vital for appropriate risk assessment, as many nongenotoxic carcinogens (including PB) are rodent specific and will not cause the same effects in humans. Liver tumor promotion in rodents by PB is dependent on the activation of the transcription factor CAR, as CAR activation has been demonstrated and CAR knockouts are not sensitive to these pleiotropic effects, namely, gene expression changes, including the induction of Cyp2b isoforms (Cyp2b10 in mice and Cyp2b2 in rats), hepatocellular hypertrophy, increased hepatocyte proliferation, and ultimately the appearance of adenomas and carcinomas [74–76]. In addition to CAR, PB also activates the pregnane X receptor (PXR) [77], which has overlapping functions with CAR to regulate xenobiotic metabolism and detoxification in liver [78]. In recent publications, Thomson, Lempiainen, and coworkers elegantly investigated the temporal sequence of PB gene regulation in mice using a battery of molecular biology techniques assessing not only gene expression patterns but also protein expression, ncRNA species, DNA methylation profiles, and SNPs [79,80]. In their work, they show that early PB-induced molecular responses include staged epigenetic perturbations, increased expression levels of Cy2b10, and the progressive increase in hepatic expression of long noncoding RNAs (lncRNAs) and miRNAs. The majority of the induced miRNAs are transcribed from the Dlk1-Dio3 imprinted genomic cluster, which has been identified as a hallmark of

pluripotency and proliferation. In addition to the miRNAs, the lncRNAs Meg3 and Rian from the maternal allele were also induced due to the treatment with PB. The Dlk1-Dio3 locus has further been associated with hepatic carcinogenesis as overexpression of the transcription factor HNF4α in human hepatocarcinoma cells resulted *in vitro* in elevated levels of miRNAs from the miR-379–656 cluster, which is located in the DLK1-DIO3 locus on human chromosome 14q32 [81]. Thus, results from epigenetic perturbations, gene expression, ncRNA expression, and SNP analysis for allele differentiation gave a clear indication of the complex molecular events of the CAR/PXR-dependent mechanism.

Besides the clear involvement in carcinogenesis, miRNAs are involved in processes related to toxicity and acute liver injury. This finding is particularly interesting since miRNAs are secreted, detectable in circulation, and highly conserved across species. This makes them extremely attractive translational, noninvasive biomarkers. In particular for the liver, the association between levels of circulating miR-122 and liver damage has been well established. This association has been observed in rodents and in humans. Elevation of miR-122 in serum in response to liver injury was first observed in mice exposed to hepatotoxic doses of cyclooxygenase inhibitor and analgesic drug acetaminophen. Moreover, serum miR-122 was significantly higher in patients presenting with acute liver injury caused by exposure with high doses of acetaminophen as well as by different etiologies. In addition, patients exposed to nontoxic levels of acetaminophen do not show increased miR-122 in the circulation. Thus, these results indicate that the increase of miR-122 is associated with hepatocellular toxicity rather than with exposure to acetaminophen. Also, the fact that mice and humans and also rats show increased miR-122 after exposure to hepatotoxins makes this noncoding RNA a putative sensitive, tissue-specific, and translational biomarker that might outperform the traditionally measured transaminases [61]. Additional investigations are necessary to improve our understanding of the causal relationship between increased miRNAs in circulation and tissue-specific events. The mechanisms by which miRNAs are specifically released from injured tissues need to be established as well as the dynamics of the signal upon damage progression and recovery.

8.4 Summary and Outlook

As described in this chapter, there is growing evidence on the relationship of genetic and genomic changes in the development of toxicological side effects. Also, interindividual and interspecies differences have a profound impact on the efficacy of medicines. Thus, the genetic background has a major impact on drug efficacy and toxicity, and therefore in safety assessment (Figure 8.2). In addition, the acknowledged relevance of epigenetic mechanisms in toxicity further promotes the study of genetic, genomic, and epigenetic changes in a concerted manner. This is only attainable by combining state-of-the-art technologies such

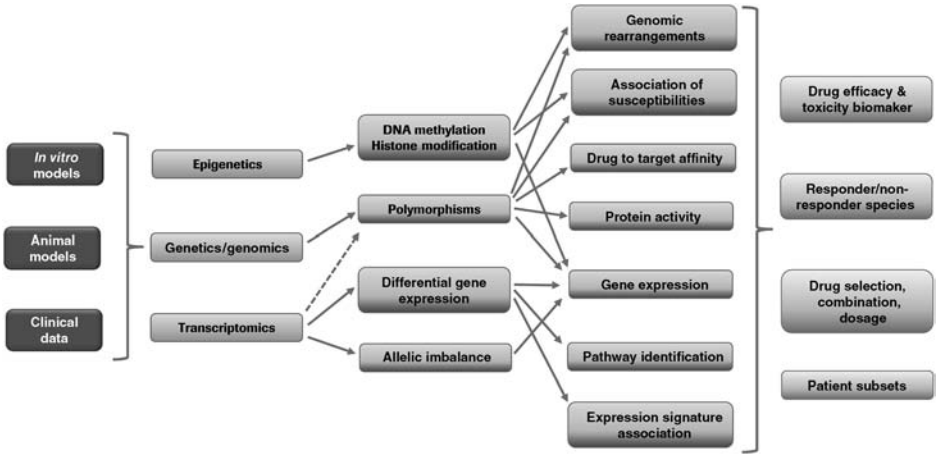


Figure 8.2 Impact of genetics, genomics, and epigenetics in toxicology. The multifaceted effects of epigenetics, DNA polymorphisms, and differential gene expression on the prediction of drug efficacy, safety/toxicity, and the susceptibility of side effects.

as advanced sequencing approaches, expression measurements, DNA methylation patterns, and imaging tools. The general trend is to employ high-throughput, information-rich technologies to assess direct effects on specific genes or pathways and clinically relevant outcome. Currently, the identification of pathways of toxicological relevance and the integration and interpretation of the data remain the main challenge to improve safety assessment based on molecular biology tools.

References

1 Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K., and Gerstein, M.B. (2011) The real cost of sequencing: higher than you think! *Genome Biol.*, **12**, 125.

2 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

3 Margulies, M., Egholm, M., Altman, W.E. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

4 Heiner, C.R., Hunkapiller, K.L., Chen, S. M., Glass, J.I., and Chen, E.Y. (1998) Sequencing multimegabase-template DNA with BigDye terminator chemistry. *Genome Res.*, **8**, 557–561.

5 Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. *Nat. Biotechnol.*, **26**, 1135–1145.

6 Rothberg, J.M., Hinz, W., Rearick, T.M. *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475**, 348–352.

7 Bragg, L.M., Stone, G., Butler, M.K., Hugenholtz, P., and Tyson, G.W. (2013) Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.*, **9**, e1003031.

8 Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.*, **8**, R143.

9 Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O'Toole, P.W. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.*, **38**, e200.

10 Fedurco, M., Romieu, A., Williams, S., Lawrence, I., and Turcatti, G. (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.*, **34**, e22.

11 Bentley, D.R., Balasubramanian, S., Swerdlow, H.P. *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.

12 Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.*, **36**, e105.

13 Hoffmann, S., Otto, C., Kurtz, S., Sharma, C.M., Khaitovich, P., Vogel, J., Stadler, P.F., and Hackermuller, J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.*, **5**, e1000502.

14 Minoche, A.E., Dohm, J.C., and Himmelbauer, H. (2011) Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.*, **12**, R112.

15 Kircher, M., Stenzel, U., and Kelso, J. (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.*, **10**, R83.

16 Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., Sidow, A., Fire, A., and Johnson, S.M. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.*, **18** 1051–1063.

17 ten Bosch, J.R. and Grody, W.W. (2008) Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J. Mol. Diagn.*, **10**, 484–492.

18 Eid, J., Fehr, A., Gray, J. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–138.

19 Fang, G., Munera, D., Friedman, D.I. *et al.* (2012) Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.*, **30**, 1232–1239.

20 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.*, **11**, 31–46.

21 Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.

22 Umer, M. and Herceg, Z. (2013) Deciphering the epigenetic code: an overview of DNA methylation analysis methods. *Antioxid. Redox Signal.*, **18**, 1972–1986.

23 Wu, H. and Zhang, Y. (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, **156**, 45–68.

24 Collins, F.S., Morgan, M., and Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, **300**, 286–290.

25 Waterston, R.H., Lindblad-Toh, K., Birney, E. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.

26 Gibbs, R.A., Weinstock, G.M., Metzker, M.L. *et al.* (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, **428**, 493–521.

27 Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**, 803–819.

28 Gibbs, R.A., Rogers, J., Katze, M.G. *et al.* (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**, 222–234.

29 Flicek, P., Aken, B.L., Ballester, B. *et al.* (2010) Ensembl's 10th year. *Nucleic Acids Res.*, **38**, D557–D562.

30 Wheeler, D.A., Srinivasan, M., Egholm, M. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–876.

31 Wadman, M. (2008) James Watson's genome sequenced at high speed. *Nature*, **452**, 788.

32 Ebeling, M., Kung, E., See, A., Broger, C., Steiner, G., Berrera, M., Heckel, T., Iniguez, L., Albert, T., Schmucki, R., Biller, H., Singer, T., and Certa, U. (2011) Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment. *Genome Res.*, **21** 1746–1756.

33 Yan, G., Zhang, G., Fang, X. *et al.* (2011) Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat. Biotechnol.*, **29**, 1019–1023.

34 Vamathevan, J.J., Hall, M.D., Hasan, S. *et al.* (2013) Minipig and beagle animal model genomes aid species selection in pharmaceutical discovery and development. *Toxicol. Appl. Pharmacol.*, **270**, 149–157.

35 Searls, D.B. (2003) Pharmacophylogenomics: genes, evolution and drug targets. *Nat. Rev. Drug Discov.*, **2**, 613–623.

36 Vallender, E.J., Xie, Z., Westmoreland, S. V., and Miller, G.M. (2010) Functional evolution of the trace amine associated receptors in mammals and the loss of TAAR1 in dogs. *BMC Evol. Biol.*, **10**, 51.

37 Sun, Y.V., Boverhof, D.R., Burgoon, L.D., Fielden, M.R., and Zacharewski, T.R. (2004) Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences. *Nucleic Acids Res.*, **32**, 4512–4523.

38 Daly, A.K. (2010) Genome-wide association studies in pharmacogenomics. *Nat. Rev. Genet.*, **11**, 241–246.

39 Karlsson, E.K. and Lindblad-Toh, K. (2008) Leader of the pack: gene mapping in dogs and other model organisms. *Nat. Rev. Genet.*, **9**, 713–725.

40 Wilbe, M., Jokinen, P., Truve, K., Seppala, E.H., Karlsson, E.K., Biagi, T., Hughes, A., Bannasch, D., Andersson, G., Hansson-Hamlin, H., Lohi, H., and Lindblad-Toh, K. (2010) Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nat. Genet.*, **42** 250–254.

41 Stevison, L.S. and Kohn, M.H. (2008) Determining genetic background in captive stocks of cynomolgus macaques (*Macaca fascicularis*). *J. Med. Primatol.*, **37**, 311–317.

42 Street, S.L., Kyes, R.C., Grant, R., and Ferguson, B. (2007) Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques. *BMC Genomics*, **8**, 480.

43 Saito, Y., Naruse, T.K., Akari, H., Matano, T., and Kimura, A. (2012) Diversity of MHC class I haplotypes in cynomolgus macaques. *Immunogenetics*, **64**, 131–141.

44 Saito, A., Kono, K., Nomaguchi, M., Yasutomi, Y., Adachi, A., Shioda, T., Akari, H., and Nakayama, E.E. (2012) Geographical, genetic and functional diversity of antiretroviral host factor TRIMCyp in cynomolgus macaque (*Macaca fascicularis*). *J. Gen. Virol.*, **93**, 594–602.

45 Chamanza, R., Marxfeld, H.A., Blanco, A. I., Naylor, S.W., and Bradley, A.E. (2010) Incidences and range of spontaneous findings in control cynomolgus monkeys (*Macaca fascicularis*) used in toxicity studies. *Toxicol. Pathol.*, **38**, 642–657.

46 Vidal, J.D., Drobatz, L.S., Holliday, D.F., Geiger, L.E., and Thomas, H.C. (2010) Spontaneous findings in the heart of Mauritian-origin cynomolgus macaques (*Macaca fascicularis*). *Toxicol. Pathol.*, **38**, 297–302.

47 Drevon-Gaillot, E., Perron-Lepage, M.F., Clement, C., and Burnett, R. (2006) A review of background findings in cynomolgus monkeys (*Macaca fascicularis*) from three different geographical origins. *Exp. Toxicol. Pathol.*, **58**, 77–88.

48 Liu, Y.Y., Wang, Z., Thomas, J., Goodwin, K.J., Stavrou, S., and Neville, D.M., Jr. (2007) Polymorphisms of CD3epsilon in cynomolgus and rhesus monkeys and their relevance to anti-CD3 antibodies and immunotoxins. *Immunol. Cell Biol.*, **85**, 357–362.

49 Ingelman-Sundberg, M., Sim, S.C., Gomez, A., and Rodriguez-Antona, C. (2007) Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.*, **116**, 496–526.

50 Johansson, I. and Ingelman-Sundberg, M. (2011) Genetic polymorphism and toxicology – with emphasis on cytochrome P450. *Toxicol. Sci.*, **120**, 1–13.

51 Evans, W.E. and Relling, M.V. (1999) Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*, **286**, 487–491.

52 Gatanaga, H., Hayashida, T., Tsuchiya, K. *et al.* (2007) Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 *6 and *26. *Clin. Infect. Dis.*, **45**, 1230–1237.

53 Sibbing, D., Gebhard, D., Koch, W., Braun, S., Stegherr, J., Morath, T., Von Beckerath, N., Mehilli, J., Schomig, A., Schuster, T., and Kastrati, A. (2010) Isolated and interactive impact of common CYP2C19 genetic variants on the antiplatelet effect of chronic clopidogrel therapy. *J. Thromb. Haemost.*, **8**, 1685–1693.

54 Wang, L., McLeod, H.L., and Weinshilboum, R.M. (2011) Genomics and drug response. *N. Engl. J. Med.*, **364**, 1144–1153.

55 Daly, A.K. (2013) Optimal dosing of warfarin and other coumarin anticoagulants: the role of genetic polymorphisms. *Arch. Toxicol.*, **87**, 407–420.

56 Hicks, J.K., Swen, J.J., Thorn, C.F., Sangkuhl, K., Kharasch, E.D., Ellingrod, V.L., Skaar, T.C., Muller, D.J., Gaedigk, A., and Stingl, J.C. (2013) Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. *Clin. Pharmacol. Ther.*, **93**, 402–408.

57 Marques, S.C. and Ikediobi, O.N. (2010) The clinical application of UGT1A1 pharmacogenetic testing: gene–environment interactions. *Hum. Genomics*, **4**, 238–249.

58 Lee, J.S., Wang, J., Martin, M., Germer, S., Kenwright, A., Benayed, R., Spleiss, O., Platt, A., Pilson, R., Hemmings, A., Weinblatt, M.E., Kaplowitz, N., and Krasnow, J. (2011) Genetic variation in UGT1A1 typical of Gilbert syndrome is associated with unconjugated hyperbilirubinemia in patients receiving tocilizumab. *Pharmacogenet. Genomics*, **21** 365–374.

59 Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M., and Collins, R. (2008) SLCO1B1 variants and statin-induced myopathy – a genomewide study. *N. Engl. J. Med.*, **359**, 789–799.

60 Di Stasi, S.L., MacLeod, T.D., Winters, J.D., and Binder-Macleod, S.A. (2010) Effects of statins on skeletal muscle: a perspective for physical therapists. *Phys. Ther.*, **90**, 1530–1542.

61 Al-Dabbagh, S.G., Idle, J.R., and Smith, R.L. (1981) Animal modelling of human polymorphic drug oxidation – the metabolism of debrisoquine and phenacetin in rat inbred strains. *J. Pharm. Pharmacol.*, **33**, 161–164.

62 Sakai, N. and Ishizuka, M. (2009) Impact of rat P450 genetic polymorphism on diazepam metabolism. *Expert Opin. Drug Metab. Toxicol.*, **5**, 1421–1433.

63 Kamimura, H. (2006) Genetic polymorphism of cytochrome P450s in beagles: possible influence of CYP1A2 deficiency on toxicological evaluations. *Arch. Toxicol.*, **80**, 732–738.

64 Emoto, C., Yoda, N., Uno, Y., Iwasaki, K., Umehara, K., Kashiyama, E., and Yamazaki, H. (2013) Comparison of P450 enzymes between cynomolgus monkeys and humans: P450 identities, protein contents, kinetic parameters, and potential for inhibitory profiles. *Curr. Drug Metab.*, **14**, 239–252.

65 Jacqz, E., Billante, C., Moysan, F., and Mathieu, H. (1988) The non-human primate: a possible model for human genetically determined polymorphisms in oxidative drug metabolism. *Mol. Pharmacol.*, **34**, 215–217.

66 Atanur, S.S., Diaz, A.G., Maratou, K. *et al.* (2013) Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell*, **154**, 691–703.

67 Boitier, E., Amberg, A., Barbie, V. *et al.* (2011) A comparative integrated transcript analysis and functional characterization of differential mechanisms for induction of liver hypertrophy in the rat. *Toxicol. Appl. Pharmacol.*, **252**, 85–96.

68 Hoflack, J.C., Mueller, L., Fowler, S., Braendli-Baiocco, A., Flint, N., Kuhlmann, O., Singer, T., and Roth, A. (2012) Monitoring Cyp2b10 mRNA expression at cessation of 2-year carcinogenesis bioassay in mouse liver provides evidence for a carcinogenic mechanism devoid of human relevance: the dalcetrapib experience. *Toxicol. Appl. Pharmacol.*, **259**, 355–365.

69 Roth, A., Boess, F., Landes, C., Steiner, G., Freichel, C., Plancher, J.M., Raab, S., de Vera Mudry, C., Weiser, T., and Suter, L. (2011) Gene expression-based *in vivo* and *in vitro* prediction of liver toxicity allows compound selection at an early stage of drug development. *J. Biochem. Mol. Toxicol.*, **25**, 183–194.

70 Szyf, M. (2011) The implications of DNA methylation for toxicology: toward toxicomethylomics, the toxicology of DNA methylation. *Toxicol. Sci.*, **120**, 235–255.

71 Tajbakhsh, J. (2011) DNA methylation topology: potential of a chromatin landmark for epigenetic drug toxicology. *Epigenomics*, **3**, 761–770.

72 Liu, H., Zhou, Y., Boggs, S.E., Belinsky, S.A., and Liu, J. (2007) Cigarette smoke induces demethylation of prometastatic oncogene synuclein-gamma in lung cancer cells by downregulation of DNMT3B. *Oncogene*, **26**, 5900–5910.

73 Whysner, J., Ross, P.M., and Williams, G.M. (1996) Phenobarbital mechanistic data and risk assessment: enzyme induction, enhanced cell proliferation, and tumor promotion. *Pharmacol. Ther.*, **71**, 153–191.

74 Kawamoto, T., Sueyoshi, T., Zelko, I., Moore, R., Washburn, K., and Negishi, M. (1999) Phenobarbital-responsive nuclear translocation of the receptor CAR in induction of the CYP2B gene. *Mol. Cell. Biol.*, **19**, 6318–6322.

75 Ross, J., Plummer, S.M., Rode, A., Scheer, N., Bower, C.C., Vogel, O., Henderson, C.J., Wolf, C.R., and Elcombe, C.R. (2010) Human constitutive androstane receptor (CAR) and pregnane X receptor (PXR) support the hypertrophic but not the hyperplastic response to the murine nongenotoxic hepatocarcinogens phenobarbital and chlordane *in vivo*. *Toxicol. Sci.*, **116**, 452–466.

76 Phillips, J.M., Burgoon, L.D., and Goodman, J.I. (2009) The constitutive active/androstane receptor facilitates unique phenobarbital-induced expression changes of genes involved in key pathways in precancerous liver and liver tumors. *Toxicol. Sci.*, **110**, 319–333.

77 Lehmann, J.M., McKee, D.D., Watson, M.A., Willson, T.M., Moore, J.T., and Kliewer, S.A. (1998) The human orphan nuclear receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest.*, **102**, 1016–1023.

78 Tolson, A.H. and Wang, H. (2010) Regulation of drug-metabolizing enzymes by xenobiotic receptors: PXR and CAR. *Adv. Drug Deliv. Rev.*, **62**, 1238–1249.

79 Lempiainen, H., Couttet, P., Bolognani, F. *et al.* (2013) Identification of Dlk1-Dio3 imprinted gene cluster noncoding RNAs as novel candidate biomarkers for liver tumor promotion. *Toxicol. Sci.*, **131**, 375–386.

80 Thomson, J.P., Hunter, J.M., Lempiainen, H., Muller, A., Terranova, R., Moggs, J.G., and Meehan, R.R. (2013) Dynamic changes in 5-hydroxymethylation signatures underpin early and late events in drug exposed liver. *Nucleic Acids Res.*, **41**, 5639–5654.

81 Yin, C., Wang, P.Q., Xu, W.P., Yang, Y., Zhang, Q., Ning, B.F., Zhang, P.P., Zhou, W.P., Xie, W.F., Chen, W.S., and Zhang, X. (2013) Hepatocyte nuclear factor-4alpha reverses malignancy of hepatocellular carcinoma through regulating miR-134 in the DLK1-DIO3 region. *Hepatology*, **58**, 1964–1976.

9
Biomarkers as Tools for Predictive Safety Assessment: Novel Markers of Drug-Induced Kidney Injury

Angela Mally

The paramount aim of predictive toxicology in drug discovery and development is to assess the safety of drugs for use in humans. Although the concept of predictive toxicology has always been part of the drug development process, the past decade has seen an enormous revival of the field mainly for two reasons: first, it has been recognized that drug safety issues have replaced unfavorable pharmacokinetics as a leading cause of drug attrition [1]; second, significant advances in science and availability of new technologies such as genomics, proteomics, and metabolomics have opened up unprecedented opportunities to study drug toxicity and to utilize this knowledge for improved safety assessment.

Biomarkers, which can be defined as "physiological or molecular measurements that indicate or predict the outcome of testing in animals or of a clinical endpoint in therapeutic trials" [2], are recognized as valuable tools that can be applied at all stages of drug development to identify drugs with unacceptable levels of toxicity earlier or with greater sensitivity or specificity than traditional approaches.

Thus, considerable efforts are being invested to identify novel biomarkers of toxicity and to develop tests based on biomarkers that can be used to diagnose and monitor drug-induced injury in major target organs, for example, the kidney, liver, and cardiovascular system, and to bridge between species and *in vitro*/*in vivo* systems. Thanks to these efforts, significant progress has been made particularly in the field of biomarkers of nephrotoxicity. This chapter will use drug-induced kidney injury as an example to illustrate the value and limitations of novel biomarkers by providing an overview of their (patho)physiological role and mechanistic link to toxicity, performance in experimental toxicity studies, approval by regulatory bodies, and application to preclinical decision making.

9.1
Need and Search for Novel Biomarkers of Kidney Injury

Due to the high blood flow rate, which causes delivery of high concentrations of drugs to the kidney, and the presence of active transporters and drug metabolizing enzymes, which enable uptake and bioactivation of xenobiotics, the kidney

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

and in particular the proximal tubule epithelium presents one of the key target sites of drug-induced injury.

Traditional markers of renal function such as serum creatinine and blood urea nitrogen (BUN) are inherently insensitive as they indicate kidney injury only when large amounts of the renal epithelium have been lost. Advances in omics technologies, in particular toxicogenomics and proteomics, have enabled the identification of genes, proteins, and more recently also miRNAs that are altered in response to kidney injury and may present more effective indicators of a drug's potential to cause nephrotoxicity than the traditional functional parameters.

Predictive biomarkers that can be measured noninvasively, that is, in urine, are recognized as particularly valuable in drug development, since they may be used in preclinical and clinical studies to continuously monitor target organ injury in real time without the need of tissue biopsies. Thus, most research efforts have been dedicated to the identification, development, and qualification of urinary biomarkers of kidney injury. In principle, increased urinary excretion of a marker in response to injury can reflect altered renal handling of substances, including low molecular weight proteins produced at other sites, due to functional impairment of tubule cells, leakage from injured tubule cells, or increased formation, for example, through upregulation of gene expression, and subsequent active or passive release into urine (Figure 9.1).

9.2 Urinary Biomarkers of Drug-Induced Kidney Injury

9.2.1 Structure and Function of Novel Urinary Biomarkers

9.2.1.1 Kidney Injury Molecule-1

Perhaps the most widely studied urinary kidney biomarker is kidney injury molecule-1 (KIM-1) encoded by the gene Havr1 (hepatitis A virus cellular receptor 1). KIM-1 is a 104 kDa type 1 transmembrane glycoprotein that is expressed at low levels at the apical membrane of proximal tubule cells but is markedly induced in response to proximal tubule injury. It contains an N-terminal T-cell immunoglobulin domain, a mucin domain, a single transmembrane domain, and a C-terminal cytoplasmic domain with a tyrosine kinase phosphorylation motif [3].

The functional roles of KIM-1 in tissue injury and repair are not fully understood. Like other members of the T-cell immunoglobulin mucin (TIM) family, KIM-1 contains a phosphatidylserine (PS) binding pocket within its N-terminal immunoglobulin domain. This allows proximal tubule cells expressing KIM-1 to recognize PS exposed on the outer membrane of apoptotic cells and to internalize apoptotic bodies and cell debris [4].

Upregulation of KIM-1 thus confers phagocytic capacity on proximal tubule cells and appears to be an important mechanism to prevent tubular obstruction, limit inflammation, and facilitate tissue regeneration through clearance of cell

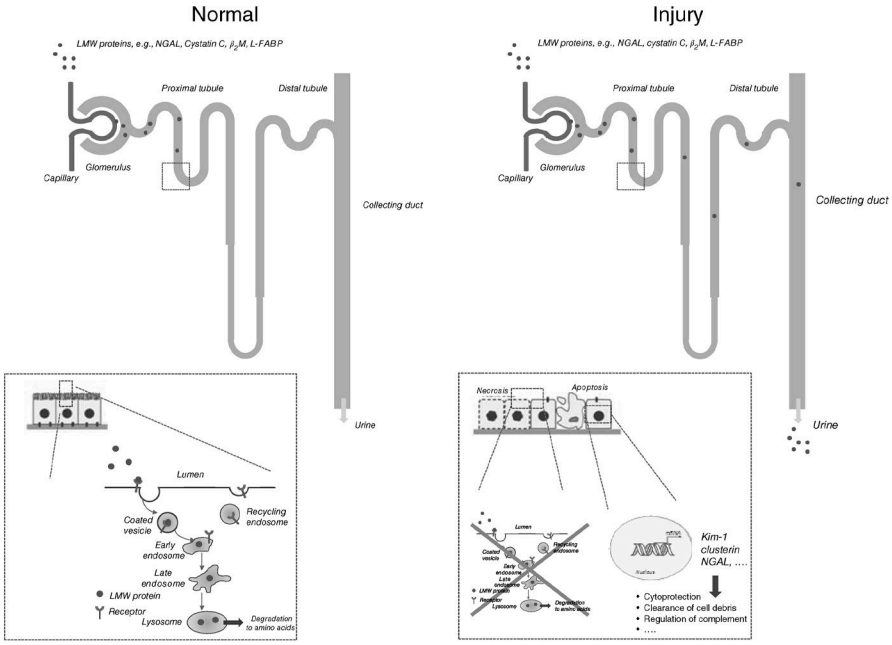


Figure 9.1 Urinary excretion of protein-based biomarkers due to altered renal handling of low molecular weight proteins under physiological and pathophysiological conditions and upregulation of gene expression in response to proximal tubule injury.

debris. Release of KIM-1 from injured proximal tubule cells involves shedding of its 90 kDa ectodomain, which is excreted in urine and can be measured as a noninvasive marker of kidney injury. Shedding of soluble KIM-1 appears to be mediated by metalloproteinases [5], but the precise signaling pathways that regulate KIM-1 release remain to be elucidated. Studies in proximal tubule cell culture models demonstrate that shedding of constitutive KIM-1 is stimulated by the proinflammatory cytokine tumor necrosis factor α (TNFα), human serum albumin, and by activation of extracellular signal-related kinase (ERK) and p38 mitogen-activated protein kinase (MAPK) [5,6].

9.2.1.2 Clusterin

Clusterin, also known as apolipoprotein J, testosterone-repressed prostate message-2, or sulfated glycoprotein-2, was originally identified as a protein causing clustering of Sertoli cells [7]. It is a secreted 75–80 kDa heterodimeric glycoprotein that is constitutively expressed in a wide range of tissues and is present in body fluids such as plasma, milk, urine, and semen [8].

It is composed of an α- and a β-chain that are held together by five disulfide bonds. Following dimer assembly and glycosylation in the Golgi apparatus, clusterin is secreted from secretory vesicles. In addition to the fairly well-characterized secreted form, truncated intracellular forms of clusterin resulting from alternative splicing that are targeted to the nucleus have been reported [9,10].

The functions of clusterin are diverse. Secreted clusterin is thought to act as a constitutive cytoprotective extracellular chaperone that forms complexes with a range of binding partners, including lipids, β-amyloid, immunoglobulins, components of the complement system, heparin, and leptin [8]. It has been implicated in cytoprotection at fluid–tissue boundaries, hormone and lipid transport, membrane recycling, and regulation of complement-mediated cell lysis. Clusterin is also involved in the regulation of cell–cell and cell–matrix interactions and thus contributes to tissue remodeling in response to injury. In contrast to the predominantly cytoprotective role of secreted clusterin, the truncated nuclear form of clusterin has been shown to act as a death signal and to trigger apoptosis [9].

Given the plethora of functions ascribed to clusterin, it is considered to play important roles in nearly all fundamental biological processes, including development and reproduction as well as during injury and diseases. Clusterin expression is induced in response to oxidative, thermal, or mechanical stress and in a range of pathological conditions, including neurodegeneration, arteriosclerosis, myocardial infarction, cancer, and acute and chronic kidney diseases such as glomerulonephritis, lupus-like nephritis, obstructive nephropathy, polycystic kidney disease, kidney tumors, ischemia–reperfusion injury, and acute and chronic nephrotoxicity [8,11–18].

In acute and chronic kidney diseases, upregulation of renal clusterin expression and secretion into the tubule lumen enable detection of clusterin in urine as a urinary biomarker that correlates with the severity of kidney injury [19]. Concentrations of clusterin in the systemic circulation may increase in response to damage to other organs, for example, in drug-induced liver injury [20].

Importantly, clusterin may be cleaved into its α- and β-chains, which may be able to pass the glomerulus [21], resulting in increased urinary concentrations of clusterin even in the absence of renal injury [20].

9.2.1.3 Cystatin C

Cystatin C is a protein expressed by virtually all nucleated cells. It is secreted into the extracellular space where it functions as an inhibitor of cysteine proteases of the cathepsin family. Due to its low molecular weight (13 kDa), cystatin C is freely filtered in the glomerulus, reabsorbed by proximal tubule cells via megalin-mediated endocytosis, and degraded in lysosomes [22].

An increase in the concentrations of cystatin C in serum reflects reduced glomerular filtration and thus serum cystatin C is increasingly being used in the clinical setting as an alternative to serum creatinine to estimate glomerular filtration rates [23]. In contrast, tubular dysfunction or injury to proximal tubule cells impairs uptake and degradation of cystatin C, resulting in increased excretion of cystatin C in urine.

Although cystatin C is not mechanistically linked to kidney injury, increased urinary cystatin C due to altered renal handling of cystatin serves as a sensitive marker of tubular dysfunction and toxicity.

9.2.1.4 β_2-Microglobulin

β_2-Microglobulin is a low molecular weight protein (11.8 kDa) that is expressed by all nucleated cells. It is a component of the major histocompatibility complex (MHC) class I, which functions to display peptides generated from degradation of cytosolic proteins to cytotoxic T cells. β_2-Microglobulin is noncovalently linked to the α-chain of MHC I, which contains a plasma membrane-spanning domain, and appears to be necessary for localization of MHC I at the cell surface. It is shed from the surface of nucleated cells and can be found in plasma and other body fluids. Like cystatin C and other low molecular weight proteins, β_2-microglobulin is freely filtered in the glomeruli and reabsorbed and degraded by proximal tubule cells [24]. Thus, levels of β_2-microglobulin in urine are normally low. Tubular dysfunction or injury, however, results in reduced absorption of low molecular weight proteins and thus increased urinary excretion of β_2-microglobulin.

9.2.1.5 Liver-Type Fatty Acid Binding Protein

Liver-type fatty acid binding protein (L-FABP) is a 14.4 kDa cytoplasmic protein that belongs to the calycin superfamily. It forms a β-barrel structure that facilitates high-affinity binding and transport of hydrophobic ligands, including long-chain and very long-chain fatty acids, fatty acyl-CoAs, eicosanoids, cholesterol, and bile acids [25].

Its functions include cellular uptake and intracellular transport of fatty acids to mitochondria and peroxisomes for fatty acid oxidation [25]. L-FABP has also been implicated in the transcriptional regulation of lipid metabolism [26]. L-FABP was shown to exert cytoprotective function against oxidative stress [27,28].

Although its name indicates that it is synthesized in the liver, L-FABP may also be expressed in other tissues, including the proximal tubule epithelium. Circulating L-FABP released from cells is filtered by glomeruli, taken up into proximal tubule cells via megalin-mediated endocytosis, and degraded to amino acids in lysosomes [29]. Increased excretion of L-FABP due to altered renal handling can thus be used as an indicator of proximal tubule dysfunction. In contrast to some of the other low molecular weight proteins that serve as markers of kindey injury, L-FABP also appears to be mechanistically linked to tubular injury as L-FABP gene expression may be upregulated in proximal tubule cells in response to stress and may protect against renal injury [30–32].

9.2.1.6 Neutrophil Gelatinase-Associated Lipocalin

Neutrophil gelatinase-associated lipocalin (NGAL, lipocalin-2, LCN2) is a 25 kDa protein that belongs to the lipocalin superfamily, which shares a structural element that allows binding and transport of low molecular weight hydrophobic ligands. NGAL was initially discovered in activated neutrophils as a novel protein covalently bound to human gelatinase/matrix metalloproteinase-9, hence the name [33]. It is an acute-phase protein that is expressed not only in neutrophils but also in a wide range of cells and tissues, including epithelial cells in the colon, bile ducts, lung, and kidney, in response to cellular and inflammatory stress [34].

Its functions are not completely understood. NGAL participates in host innate immune defense by binding bacterial siderophores to limit bacterial iron acquisition. Expression of NGAL at possible sites of bacterial entry and rapid induction of NGAL transcription may thus serve as a first line of defense against bacterial infections. There is some evidence to suggest that NGAL may also play a role in cell proliferation and protection from apoptosis [34]. Like other low molecular weight proteins, NGAL circulating in plasma passes the glomerulus and is reabsorbed into proximal tubule cells by receptor-mediated endocytosis. In proximal tubule cells, NGAL localizes to lysosomes, where it is catabolized to amino acids. Thus, only low baseline levels of NGAL are normally detectable in urine.

Marked induction of NGAL during systemic inflammation or tissue injury and release of high amounts of NGAL into the systemic circulation may, however, overwhelm the capacity for tubular reabsorption and result in a state of overflow, where NGAL is excreted in urine. In the kidney, NGAL may be rapidly upregulated at sites of injury and secreted into urine and plasma via the apical and basolateral membrane of epithelial cells, leading to increased urine and plasma concentrations of NGAL. Impaired proximal tubule reabsorption due to damage to the proximal tubule epithelium may further augment urinary NGAL excretion.

9.2.1.7 Others

In addition to the markers of nephrotoxicity summarized above, a number of other urinary proteins have been put forward as potential markers of proximal tubule injury. These include vanin-1, an epithelial glycosylphosphatidylinositol-

anchored pantetheinase that appears to play a role in oxidative stress response and may leak from kidney tubule cells [35], glutathione *S*-transferase alpha (GST-α), a cytosolic phase II drug metabolizing enzyme that is passively released from injured proximal tubule cells into urine, vascular endothelial growth factor (VEGF), calbindin, albumin, trefoil factor 3 (TFF3), fibrinogen [36], osteopontin, and tissue inhibitor of metalloproteinases-1 (*Timp-1*), which are also regulated in the kidney in response to proximal tubule damage and are outlined below (see Section 9.3).

9.2.2 Experimental and Clinical Support for the Use of Novel Urinary Biomarkers for the Detection and Prediction of Acute Kidney Injury

9.2.2.1 Performance of Novel Urinary Biomarkers in Preclinical Models of Renal Injury

Since urinary proteins such as KIM-1 first came into light as novel biomarkers of nephrotoxicity about 10 years ago [37], there has been an ever-increasing number of experimental studies in laboratory animals addressing the performance of individual urinary markers or biomarker panels in detecting (drug-induced) kidney injury.

KIM-1, for instance, has been shown to be increased in rat urine in response to treatment with a range of nephrotoxins, including cisplatin [37–47], gentamicin [17,35,40,48–50], vancomycin [43], ochratoxin A [17,51], melamine and cyanuric acid [52], cadmium [53], mercury [48], chromium [48], paraquat [54], and glyphosate-based herbicide [55]. Although the experimental study designs and degree of kidney injury induced by the treatments greatly varied between studies, urinary excretion of KIM-1 was generally reported to be an early and sensitive indicator of nephrotoxicity that correlated well with the onset, progression, and recovery from proximal tubule injury and the severity of lesions.

Frequently, receiver operating characteristic (ROC) analyses were carried out to demonstrate that KIM-1 outperforms traditional clinical chemistry such as BUN and serum creatinine in terms of sensitivity and specificity in detecting kidney injury in rats. Considering the continuously growing body of data in rats, it is, however, surprising that – with the exception of a recent study on polymyxin B nephrotoxicity across multiple species (rats, dogs, and monkeys) [56] – there are no (published) reports on the performance of KIM-1 in nonrodent preclinical species.

Similar to KIM-1, numerous authors investigated the suitability of urinary NGAL and urinary clusterin to monitor drug-induced kidney injury in rodents. Increased urinary excretion of clusterin was demonstrated in response to kidney injury induced, for example, by cisplatin [39,41,43,44,46,57–59], gentamicin [17,18,40,49,50,58], vancomycin [43], melamine and cyanuric acid [52], and lisinopril and rosuvastatin [60], while the ability of urinary NGAL to detect kidney injury was confirmed in rats treated with cisplatin [35,39,40,43,46,47], gentamicin [17,35,40,49,50], vancomycin [43], and ochratoxin A [17]. Urinary NGAL was also found to be a sensitive biomarker of acute kidney injury in gentamicin-

and polymyxin B-induced nephrotoxicity models in dogs [56,61], demonstrating its potential value for cross-species translation. Similarly, a recent study in a population of dogs suffering from renal injury due to leishmaniasis suggests that urinary clusterin may be a suitable renal injury marker in dogs [62].

The relative performance of individual urinary biomarkers in detecting drug-induced kidney injury was assessed in a range of studies using combinations of individual assays or multiplex assays for the simultaneous analysis of biomarker panels [17,40,43,52,57–59,63,64]. From these studies, it is evident that the individual biomarker responses may vary depending on the type of nephrotoxin under investigation [17,40]. While KIM-1 was frequently found to outperform other markers in terms of sensitivity and specificity [17,40,63–65], there are also a number of examples where impaired tubular reabsorption of low molecular weight proteins such as cystatin C, NGAL, or β_2-microglobulin was an earlier indicator of drug-induced proximal tubule injury that enabled detection of the prodromal stage of toxicity [17,40]. For instance, increased urinary concentrations of cystatin C and NGAL were observed in rats treated with the aminoglycoside gentamicin before changes in urinary KIM-1 were evident [17]. Similarly, NGAL was given preference over KIM-1 to screen novel kidney-sparing polypeptide-based antibiotics in rats [56].

Mechanistically, the superior performance of low molecular weight proteins compared with tubule injury proteins such as KIM-1 may be explained by competitive inhibition of receptor-mediated tubular uptake by drugs such as aminoglycosides and polymyxin analogs, which are also substrates of the megalin transporter and may displace low molecular weight proteins from renal brush border binding sites. It is therefore plausible that increased urinary excretion of low molecular weight protein markers may occur even in the absence of proximal tubule toxicity.

On the other hand, it is critical to recognize that (in contrast to KIM-1, which is specifically induced in kidney epithelial cells in response to injury) low molecular weight protein markers show some lack of specificity. Increased systemic concentrations due to production/release at other sites, for example, during systemic inflammation, drug-induced liver injury, and muscle injury, may overwhelm the capacity of tubular reabsorption and lead to increased excretion of low molecular weight proteins in urine [20,64]. Glomerular injury leading to leakage of proteins into urine may also result in increased urinary biomarker levels due to protein overload in the tubules.

9.2.2.2 Clinical Support for Novel Urinary Kidney Injury Biomarkers

Acute kidney injury, which is characterized by a rapid and sustained decrease in kidney function, presents an increasing and challenging clinical problem. Thus, there has been an enormous interest by clinicians to utilize emerging urinary kidney injury markers for the (differential) diagnosis, prognosis, and management of acute kidney injury in the clinic.

KIM-1, for instance, was shown to provide a sensitive indicator of acute kidney injury in patients suffering from acute kidney injury after cardiac surgery,

ischemia, obstructive nephropathy, acute paraquat intoxication, in renal transplant recipients, and in the critically ill [66–73].

The utility of NGAL as a novel urinary marker for the diagnosis of acute kidney injury was investigated in transplant patients [74,75], after adult and pediatric cardiac surgery [76,77], in critically ill children with septic shock [78], in neonates [79,80], in critically ill multiple trauma patients [81], and in patients treated with cisplatin [82], to name just a few. These studies generally show that measurement of urinary NGAL can improve early diagnosis of acute kidney injury in patients, leading some authors to propose NGAL as a "renal troponin" [77], in analogy to cardiac troponin, a highly sensitive and specific marker of acute myocardial injury.

However, there are few clinical data on the specificity of NGAL and other emerging markers of renal injury. There is also no consensus as to which biomarker may be the most appropriate for each purpose and how new biomarkers may be incorporated into routine practice to improve clinical outcomes. Regardless of these constraints, results from a large number of studies in patients with acute kidney injury support the potential value of novel renal injury biomarkers as translational markers for early detection of drug-induced kidney injury in clinical studies during drug development and efforts are currently underway to qualify novel kidney injury biomarkers for use in clinical trials (http://www.imi-safe-t.eu/).

9.3 Genomic Biomarkers

It is well established that cells sense and respond to cellular stress and toxicity through up- and downregulation of gene expression. It is also accepted that perturbation of pathways of toxicity can cause changes in gene expression in the target organ even in the absence of overt toxicity, for example, as an adaptive means to increase mechanisms of cell defense and maintain tissue homeostasis. Although tissue biomarkers are invasive and – in contrast to urinary markers – not amenable to the clinic in general, it has been reasoned that alterations in gene expression may encompass truly predictive markers of nephrotoxicity that precede cytotoxicity and may add to the preclinical safety toolbox. Since the introduction of toxicogenomics, substantial efforts have thus been invested to identify individual genes or signatures of genes that are deregulated in response to nephrotoxic drugs and might serve as early mechanism-based indicators of drug-induced kidney injury.

9.3.1 Individual Genes

Many of the above-mentioned protein markers had been originally detected based on genomic approaches. Transcription profiling identified KIM-1 (*Havr1*),

clusterin (*Clu*), NGAL (*Lcn2*), *Timp-1*, osteopontin (*Spp1*), vimentin (*Vim*), and heme oxygenase 1 (*Hmox-1*) as genes frequently upregulated in response to drug-induced kidney injury [83–85].

Timp-1 inhibits the proteolytic activity of matrix metalloproteinases and has been implicated in regeneration and protection from apoptosis. Vimentin is an intermediary filament that is expressed in dedifferentiated cells with a mesenchymal phenotype and is thus considered to present a marker of regeneration.

Similarly, osteopontin, which can be detected primarily in dedifferentiated tubule epithelial cells in response to injury, is thought to play a role in tissue regeneration.

Heme oxygenase-1 has cytoprotective antioxidative functions and is rapidly upregulated under conditions of oxidative and cellular stress.

Time- and dose-dependent changes in the expression of these putative biomarkers were generally found to correlate well with the severity of histopathological lesions and preceded effects on traditional clinical parameters indicative of impaired kidney function [51,86]. Induction of KIM-1 mRNA was often one of the earliest and most prominent responses observed. However, it is important to recognize that (with the exception of heme oxygenase-1) all of these markers are mechanistically linked to degeneration and regeneration, and there is no evidence to suggest that any of these markers precede tissue injury and may be truly predictive of nephrotoxicity. Dramatic upregulation of, for example, KIM-1 even in just few affected cells, which may not be noticed during histopathological evaluation of random tissue sections, may give rise to a prominent gene expression signal that enhances the ability to detect even minimal lesions. Changes in the expression of kidney injury marker genes such as KIM-1 – albeit not predictive – may thus be a more objective and quantifiable measure than routine histopathology and valuable addition to the diagnostic toolkit.

9.3.2 Biomarker Panels and Gene Signatures

Besides differential expression of individual kidney injury marker genes, it has been speculated that prediction models based on gene expression profiles induced by nephrotoxic compounds may allow identification of drugs with nephrotoxic potential earlier than routine histopathology.

Although several gene expression signatures predictive of nephrotoxicity have been reported, it is important to emphasize that the experimental study design of these early toxicogenomic studies was often such that allowed (early) diagnosis of injury based on profiling of gene expression changes associated with renal injury, but not necessarily prediction of future onset toxicity [83,84,87–89].

Not surprisingly, kidney injury marker genes such as KIM-1, clusterin, NGAL, and osteopontin were frequently part of such gene panels. The close link of the molecular changes to tissue injury is also evident from gene ontology analysis of genes contained in the signatures, which primarily include genes involved in

degenerative and regenerative processes, including inflammation, coagulation, cell growth, and tissue remodeling. A cross-study comparison of kidney gene expression profiles induced by three nephrotoxic drug candidates also revealed modulation of the complement system as a common effect associated with drug-induced kidney injury [89].

More recently, Minowa *et al.* identified a gene signature consisting of 19 probes that allowed prediction of future onset of proximal tubule injury in rats with high accuracy [90]. This multigene-based toxicogenomic model, measured 24 h after a single dose of each nephrotoxicant before histopathological changes could be detected, was able to predict repeated dose toxicity of a wide range of nephrotoxicants with a sensitivity of 93% and a selectivity of 90% [90]. In contrast to a previous diagnostic model by the same group [88], this signature was enriched with genes involved in DNA replication, cell cycle control, apoptosis, and response to chemical/oxidative stress, that is, genes that may reflect perturbation of toxicity pathways rather than overt toxicity.

9.3.3 MicroRNAs

MicroRNAs (miRNAs) are short (20–22 nucleotides in length), highly conserved noncoding RNAs involved in the complex network of gene regulation. miRNAs control gene expression at the posttranscriptional level through binding to complementary sequences in the 3′-untranslated region (3′UTR) of target mRNAs, resulting in repression of translation or mRNA.

miRNAs thus play an essential role during both normal and pathological processes, and aberrant expression of miRNAs has been linked to a wide variety of pathological conditions, including neurodegeneration, allergy, rheumatoid arthritis, autoimmune disorders, osteoporosis, diabetes, cardiovascular diseases, pulmonary diseases, and a variety of cancers. Not surprisingly, miRNAs have also been implicated in human kidney diseases such as acute kidney injury [91,92], polycystic kidney disease [93], lupus nephritis [94], diabetic nephropathy [95], and renal cell tumors [96,97].

It was recently discovered that miRNAs are contained within exosomes. Exosomes are small cell-derived extracellular vesicles that participate in intercellular communication and immune regulation through shuttling of molecules between cells. Since exosomes are present in most body fluids, including urine, there is growing interest to utilize urinary exosomes as a source for renal biomarker discovery [98,99].

In a recent study investigating the expression and excretory profile of miRNAs during acute kidney injury, miR-21, miR-155, and miR-18a were found to be among the most upregulated miRNA species in rat kidney following ischemic or toxic insult [92]. Urinary excretion of exosomal miR-21 was increased by about twofold in response to reperfusion injury, whereas decreased levels of miR-21 and miR-155 were observed in a model of gentamicin nephrotoxicity. A slight but statistically significant increase in urinary miR-21 and

decrease in urinary miR-155 were detected in patients diagnosed with acute kidney injury compared with healthy individuals, leading the authors to speculate that miR-21 and miR-155 may have potential as translational biomarkers for acute kidney injury [92].

Based on the functional analysis of predicted mRNA targets, it appears that miR-21 and miR-155 are mechanistically linked to apoptosis and cell proliferation [92]. In addition, there is evidence from the literature that miR-21 and miR-155 play a role in the regulation of inflammation [91,100,101].

In another study employing renal ischemia–reperfusion and streptozotocin (STZ)-induced renal injury in the mouse as models of acute and chronic kidney injury, urinary excretion of miR-10a and miR-30d was found to be significantly elevated in response to kidney injury [102]. The levels of miR-10a and miR-30d in urine correlated with the severity of renal lesions and outperformed BUN in detecting kidney injury [102].

Since these miRNA species are relatively enriched in kidney tissue compared with other tissues, it has been speculated that urinary miR-10a and miR-30d may serve as noninvasive, sensitive, and specific markers for the detection of kidney injury [102].

Clearly, further research is needed to understand the function, expression, and release of miRNAs during renal diseases and to further explore their potential as novel diagnostic or even predictive biomarkers of kidney injury. However, these studies highlight the utility of urinary exosomes as a valuable resource for biomarker discovery.

9.4 Qualification and Use of Novel Kidney Injury Biomarkers in Preclinical Safety Assessment

9.4.1 Biomarker Qualification and Regulatory Acceptance

Following careful analytical validation and evaluation of the performance of novel urinary biomarkers in drug toxicity studies in rats, the Predictive Safety Testing Consortium's (PSTC) Nephrotoxicity Working Group, a pharmaceutical industry public–private partnership, submitted a recommendation to the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for the use of novel urinary biomarkers of kidney injury in regulatory decision making [103].

This biomarker qualification process led to the regulatory acceptance of a set of seven urinary markers (KIM-1, albumin, total protein, β_2-microglobulin, cystatin C, clusterin, and trefoil factor 3) as early diagnostic biomarkers of drug-induced acute kidney tubular and glomerular alterations in the rat [103].

In a subsequent biomarker qualification submission initiated by the International Life Sciences Institute's Health and Environmental Sciences Institute

(ILSI/HESI), the US FDA and EMA issued approval for two additional urinary renal safety biomarkers (renal papillary antigen-1 and clusterin) for use in detecting acute drug-induced renal tubule alterations in the rat [104].

Recently, L-FABP was approved as a new tubular biomarker by the Japanese Ministry of Health, Labour and Welfare [30]. Clinical qualification of several of these novel urinary markers is ongoing.

9.4.2 Application of Novel Renal Safety Markers to Preclinical Decision Making

Despite significant scientific progress and successful regulatory qualification of urinary renal safety biomarkers, there is still uncertainty within the scientific community and drug industry as to which and how many urinary and/or genomic markers should be used and how they are best applied to drug safety assessment.

One of the few published examples of how novel urinary biomarkers can improve preclinical decision making involves the *in vivo* differentiation of chemical polymyxin analogs (a class of antibiotics essential for the treatment of life-threatening infections with Gram-negative bacteria) with varying nephrotoxic potential based on measurements of acute kidney injury biomarkers [56,63]. In a rat study, Keirstead *et al.* investigated the *in vivo* toxicity of polymyxin derivatives with different cytotoxicity profiles in HK-2 cells using novel renal safety biomarkers and found good correlation between the *in vitro* data and rat *in vivo* data, demonstrating the value of integrating novel safety markers into a screening cascade to facilitate identification of polymyxin class antibiotics with improved renal safety profiles [63].

Similarly, Burt *et al.* applied a cross-species biomarker-driven strategy for selection of polymyxin analogs with more favorable kidney toxicity profiles [56]. In this study, the performance of novel kidney biomarkers for the detection of acute kidney injury induced by polymyxin B (the parent compound of this class) was first investigated across multiple species (rats, dogs, and monkeys), whereby urinary KIM-1 and NGAL were identified as the most sensitive indicators of polymyxin nephrotoxicity. In a 2-day rat toxicity study, analysis of urinary NGAL was then used as a sensitive indicator of tubular toxicity to screen and rank novel polymyxin analogs to aid compound selection [56].

These examples not only highlight how novel biomarkers can successfully be incorporated into predictive toxicological screening approaches during early stages of drug development but also offer a strategy for the selection of the most appropriate biomarker(s) on a case-by-case basis. Since NGAL was found to be superior to other renal injury markers in detecting nephrotoxic effects of polymyxin antibiotics, it appears that routine analysis of the entire panel of novel biomarkers would create large amounts of unnecessary data and drive up the costs without further improving compound selection. In addition, conflicting results from different markers will pose a challenge for toxicologists and regulators in terms of safety assessment.

9.4.3 Technological Aspects

A wide range of commercially available assays and technology platforms are now available for the analysis of urinary and genomic renal safety biomarkers, particularly for rats as the most widely used preclinical animal species during early drug development.

Besides gene expression microarrays such as Affymetrix GeneChips or Illumina BeadArrays, these include multiplex qRT-PCR assays that allow simultaneous analysis of a panel of selected renal injury marker genes in a single run. SABiosciences Corporation, for instance, has developed the Nephrotoxicity RT^2 *Profiler*™ PCR array, a SYBR® Green qRT-PCR assay that measures up to 84 kidney marker genes in a 96- or 384-well format, whereas Althea DX4 provides a qRT-PCR assay for the detection of kidney injury based on a set of 33 genes, including KIM-1, clusterin, Hmox-1, SPP1, and Vim. Compugen Ltd offers qRT-PCR-based analysis of a combination of four biomarkers combined with a random forest classification model for early detection of drug-induced nephrotoxicity.

With the exception of a recent exploratory study that suggests that the Compugen model may have predictive value [105], there are, however, no publicly available data to show how these genomic biomarker assays perform in practice.

Urinary protein-based biomarkers can be analyzed by single or multiplex immunoassays. Several providers offer enzyme-linked immunosorbent assays (ELISAs) for the detection of single analytes in rat urine, such as the rat KIM-1 ELISA test kits available from Argutus Medical Ltd, R&D Systems, BioVendor R&D, or Abcam, the Clusterin Rat ELISA assay from BioVendor R&D, or NGAL ELISA assays from BioPorto. Recently, the H-RENA Strip™ Lateral-Flow Tests, a dipstick assay that allows rapid analysis of urinary KIM-1 within 15 min, has become available.

Multiplex immunoassays utilizing, for example, Luminex® xMAP® (Luminex) or MULTI-ARRAY® (MesoScale Discovery) technologies enable analysis of combinations of urinary biomarkers in a single run.

Fixed kidney injury biomarker panels such as the MesoScale Discovery® Kidney Injury Panels 1 (albumin, KIM-1, NGAL, osteopontin) and 2 (albumin, KIM-1, NGAL, osteopontin, α-GST, clusterin) or the Argutus AKI™ Test, which covers GST-α, GSTY1b, and RPA-1, are currently marketed.

The previously available Widescreen™ Rat Kidney Toxicity Panels 1 and 2 provided by Merck Millipore that each included five kidney markers (KIM-1, VEGF, calbindin, clusterin, and osteopontin and calbindin, clusterin, NGAL, cystatin C, and osteopontin, respectively) were recently discontinued and replaced by the Milliplex™ MAP Rat Kidney Toxicity Magnetic Bead Panels 1 and 2, which offer the customer the flexibility to select from a list of analytes and design a biomarker panel tailored to the specific needs.

For most kidney biomarkers, commercially available single or multiplex tests for analysis of human samples allow translation to the clinical situation.

In contrast, only a limited number of assays for urinary biomarker analysis in preclinically relevant species other than rats are currently available. These include the Canine kidney injury molecule 1 (Kim-1) ELISA Kit, a Clusterin Canine ELISA test (BioVendor R&D), NGAL ELISA assays for dogs, pigs, and monkeys (BioPorto), and the Milliplex^{TM} MAP Canine Kidney Toxicity Magnetic Bead Panel 1 (Merck Millipore), which covers KIM-1, clusterin, and cystatin C.

Comparative assessment of the most widely used platforms shows overall good concordance between the ability of the MesoScale Discovery and Luminex multiplex technology-based assays to detect drug-induced kidney injury [43,46,106].

However, absolute concentrations of individual biomarkers were shown to vary between platforms, demonstrating the importance of defining platform-specific baseline levels and cutoff values. Critical issues such as sampling and storage conditions that may negatively affect the stability of biomarkers in urine and hence performance of assays have received only little attention so far [107,108].

9.5 Summary and Perspectives

Scientific advances and tremendous efforts in the field of biomarker discovery and development have led to the establishment of a set of novel biomarkers that can be used to improve early detection of acute drug-induced kidney injury in preclinical safety studies.

In particular, novel protein-based biomarkers that are excreted in urine and can thus be measured noninvasively have emerged as a valuable addition to the diagnostic toolkit and several such candidates have gained regulatory acceptance subsequent to a biomarker qualification process. Most markers show good correlation with the degree of kidney injury.

Although individual novel kidney injury markers may occasionally detect prodromal stages of tubular toxicity that may not be apparent by histopathological evaluation, it is important to recognize that most of the genomic and urinary markers are mechanistically linked to tubular injury and dysfunction, and therefore present sensitive diagnostic rather than truly predictive markers.

Improved understanding of mechanisms of drug-induced kidney injury and mapping of toxicity pathways as proposed by the Toxicology in the 21st Century (Tox21) program may eventually lead to the identification of biomarkers that can predict future onset toxicity.

Advances in science and technology (e.g., microRNAs, next-generation sequencing) may also offer unique opportunities for the discovery of predictive biomarkers. Despite a wealth of data on the performance of individual biomarkers in experimental toxicity studies, there is no consensus as to how these novel biomarkers are best utilized, for example, which biomarkers or combinations of biomarkers should be included into the test battery. This is in part due to the observation that the relative performance of individual biomarkers may

depend on the compound under investigation. In practice, biomarker selection tailored to the specific problem may thus prove most advantageous and cost effective.

Efforts to translate novel biomarkers of acute kidney injury into other preclinically relevant species and humans for use in clinical trials are currently underway (http://www.imi-safe-t.eu/). Further research is needed to establish whether novel acute kidney injury biomarkers also show value for the detection/prediction of chronic nephrotoxicity and carcinogenicity.

References

1 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.*, **3**, 711–715.

2 Goodsaid, F.M. (2004) Identification and measurement of genomic biomarkers of nephrotoxicity. *J. Pharmacol. Toxicol. Methods*, **49**, 183–186.

3 Lim, A.I., Tang, S.C., Lai, K.N., and Leung, J.C. (2013) Kidney injury molecule-1: more than just an injury marker of tubular epithelial cells? *J. Cell Physiol.*, **228**, 917–924.

4 Ichimura, T., Asseldonk, E.J., Humphreys, B.D., Gunaratnam, L., Duffield, J.S., and Bonventre, J.V. (2008) Kidney injury molecule-1 is a phosphatidylserine receptor that confers a phagocytic phenotype on epithelial cells. *J. Clin. Invest.*, **118**, 1657–1668.

5 Lim, A.I., Chan, L.Y., Lai, K.N., Tang, S.C., Chow, C.W., Lam, M.F., and Leung, J.C. (2012) Distinct role of matrix metalloproteinase-3 in kidney injury molecule-1 shedding by kidney proximal tubular epithelial cells. *Int. J. Biochem. Cell Biol.*, **44**, 1040–1050.

6 Zhang, Z., Humphreys, B.D., and Bonventre, J.V. (2007) Shedding of the urinary biomarker kidney injury molecule-1 (KIM-1) is regulated by MAP kinases and juxtamembrane region. *J. Am. Soc. Nephrol.*, **18**, 2704–2714.

7 Blaschuk, O., Burdzy, K., and Fritz, I.B. (1983) Purification and characterization of a cell-aggregating factor (clusterin), the major glycoprotein in ram rete testis fluid. *J. Biol. Chem.*, **258**, 7714–7720.

8 Jones, S.E. and Jomary, C. (2002) Clusterin. *Int. J. Biochem. Cell Biol.*, **34**, 427–431.

9 Leskov, K.S., Klokov, D.Y., Li, J., Kinsella, T.J., and Boothman, D.A. (2003) Synthesis and functional analyses of nuclear clusterin, a cell death protein. *J. Biol. Chem.*, **278**, 11590–11600.

10 Reddy, K.B., Jin, G., Karode, M.C., Harmony, J.A., and Howe, P.H. (1996) Transforming growth factor beta (TGF beta)-induced nuclear localization of apolipoprotein J/clusterin in epithelial cells. *Biochemistry*, **35**, 6157–6163.

11 Chevalier, R.L. (2000) Obstructive nephropathy: lessons from cystic kidney disease. *Nephron*, **84**, 6–12.

12 Moll, S., Menoud, P.A., French, L., Sappino, A.P., Pastore, Y., Schifferli, J.A., and Izui, S. (1998) Tubular up-regulation of clusterin mRNA in murine lupus-like nephritis. *Am. J. Pathol.*, **152**, 953–962.

13 Silkensen, J.R., Schwochau, G.B., and Rosenberg, M.E. (1994) The role of clusterin in tissue injury. *Biochem. Cell Biol.*, **72**, 483–488.

14 Parczyk, K., Pilarsky, C., Rachel, U., and Koch-Brandt, C. (1994) Gp80 (clusterin; TRPM-2) mRNA level is enhanced in human renal clear cell carcinomas. *J. Cancer Res. Clin. Oncol.*, **120**, 186–188.

15 Dvergsten, J., Manivel, J.C., Correa-Rotter, R., and Rosenberg, M.E. (1994) Expression of clusterin in human renal diseases. *Kidney Int.*, **45**, 828–835.

16 Rosenberg, M.E. and Paller, M.S. (1991) Differential gene expression in the recovery from ischemic renal injury. *Kidney Int.*, **39**, 1156–1161.

17 Hoffmann, D., Fuchs, T.C., Henzler, T., Matheis, K.A., Herget, T., Dekant, W., Hewitt, P., and Mally, A. (2010) Evaluation of a urinary kidney biomarker panel in rat models of acute and subchronic nephrotoxicity. *Toxicology*, **277**, 49–58.

18 Eti, S., Cheng, C.Y., Marshall, A., and Reidenberg, M.M. (1993) Urinary clusterin in chronic nephrotoxicity in the rat. *Proc. Soc. Exp. Biol. Med.*, **202**, 487–490.

19 Hidaka, S., Kranzlin, B., Gretz, N., and Witzgall, R. (2002) Urinary clusterin levels in the rat correlate with the severity of tubular damage and may help to differentiate between glomerular and tubular injuries. *Cell Tissue Res.*, **310**, 289–296.

20 Adler, M., Hoffmann, D., Ellinger-Ziegelbauer, H., Hewitt, P., Matheis, K., Mulrane, L., Gallagher, W.M., Callanan, J.J., Suter, L., Fountoulakis, M.M., Dekant, W., and Mally, A. (2010) Assessment of candidate biomarkers of drug-induced hepatobiliary injury in preclinical toxicity studies. *Toxicol. Lett.*, **196**, 1–11.

21 Shannan, B., Seifert, M., Leskov, K., Willis, J., Boothman, D., Tilgen, W., and Reichrath, J. (2006) Challenge and promise: roles for clusterin in pathogenesis, progression and therapy of cancer. *Cell Death Differ.*, **13**, 12–19.

22 Kaseda, R., Iino, N., Hosojima, M., Takeda, T., Hosaka, K., Kobayashi, A., Yamamoto, K., Suzuki, A., Kasai, A., Suzuki, Y., Gejyo, F., and Saito, A. (2007) Megalin-mediated endocytosis of cystatin C in proximal tubule cells. *Biochem. Biophys. Res. Commun.*, **357**, 1130–1134.

23 Odutayo, A. and Cherney, D. (2012) Cystatin C and acute changes in glomerular filtration rate. *Clin. Nephrol.*, **78**, 64–75.

24 Ravnskov, U., Johansson, B.G., and Gothlin, J. (1972) Renal extraction of β_2-microglobulin. *Scand. J. Clin. Lab. Invest.*, **30**, 71–75.

25 Kamijo-Ikemori, A., Sugaya, T., and Kimura, K. (2006) Urinary fatty acid binding protein in renal disease. *Clin. Chim. Acta*, **374**, 1–7.

26 Schroeder, F., Petrescu, A.D., Huang, H., Atshaves, B.P., McIntosh, A.L., Martin, G.G., Hostetler, H.A., Vespa, A., Landrock, D., Landrock, K.K., Payne, H.R., and Kier, A.B. (2008) Role of fatty acid binding proteins and long chain fatty acids in modulating nuclear receptors and gene transcription. *Lipids*, **43**, 1–17.

27 Wang, G., Gong, Y., Anderson, J., Sun, D., Minuk, G., Roberts, M.S., and Burczynski, F.J. (2005) Antioxidative function of L-FABP in L-FABP stably transfected Chang liver cells. *Hepatology*, **42**, 871–879.

28 Yan, J., Gong, Y., She, Y.M., Wang, G., Roberts, M.S., and Burczynski, F.J. (2009) Molecular mechanism of recombinant liver fatty acid binding protein's antioxidant activity. *J. Lipid Res.*, **50**, 2445–2454.

29 Oyama, Y., Takeda, T., Hama, H., Tanuma, A., Iino, N., Sato, K., Kaseda, R., Ma, M., Yamamoto, T., Fujii, H., Kazama, J.J., Odani, S., Terada, Y., Mizuta, K., Gejyo, F., and Saito, A. (2005) Evidence for megalin-mediated proximal tubular uptake of L-FABP, a carrier of potentially nephrotoxic molecules. *Lab. Invest.*, **85**, 522–531.

30 Kamijo-Ikemori, A., Ichikawa, D., Matsui, K., Yokoyama, T., Sugaya, T., and Kimura, K. (2013) Urinary L-type fatty acid binding protein (L-FABP) as a new urinary biomarker promulgated by the Ministry of Health, Labour and Welfare in Japan. *Rinsho Byori*, **61**, 635–640.

31 Matsui, K., Kamijo-Ikemorif, A., Sugaya, T., Yasuda, T., and Kimura, K. (2011) Renal liver-type fatty acid binding protein (L-FABP) attenuates acute kidney injury in aristolochic acid nephrotoxicity. *Am. J. Pathol.*, **178**, 1021–1032.

32 Negishi, K., Noiri, E., Sugaya, T., Li, S., Megyesi, J., Nagothu, K., and Portilla, D. (2007) A role of liver fatty acid-binding protein in cisplatin-induced acute renal failure. *Kidney Int.*, **72**, 348–358.

33 Kjeldsen, L., Johnsen, A.H., Sengelov, H., and Borregaard, N. (1993) Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J. Biol. Chem.*, **268**, 10425–10432.

34 Singer, E., Marko, L., Paragas, N., Barasch, J., Dragun, D., Muller, D.N., Budde, K., and Schmidt-Ott, K.M. (2013) Neutrophil gelatinase-associated lipocalin: pathophysiology and clinical applications. *Acta Physiol.*, **207**, 663–672.

35 Hosohata, K., Ando, H., and Fujimura, A. (2012) Urinary vanin-1 as a novel biomarker for early detection of drug-induced acute kidney injury. *J. Pharmacol. Exp. Ther.*, **341**, 656–662.

36 Hoffmann, D., Bijol, V., Krishnamoorthy, A., Gonzalez, V.R., Frendl, G., Zhang, Q., Goering, P.L., Brown, R.P., Waikar, S.S., and Vaidya, V.S. (2012) Fibrinogen excretion in the urine and immunoreactivity in the kidney serves as a translational biomarker for acute kidney injury. *Am. J. Pathol.*, **181**, 818–828.

37 Ichimura, T., Hung, C.C., Yang, S.A., Stevens, J.L., and Bonventre, J.V. (2004) Kidney injury molecule-1: a tissue and urinary biomarker for nephrotoxicant-induced renal injury. *Am. J. Physiol. Renal Physiol.*, **286**, F552–F563.

38 Vaidya, V.S., Ramirez, V., Ichimura, T., Bobadilla, N.A., and Bonventre, J.V. (2006) Urinary kidney injury molecule-1: a sensitive quantitative biomarker for early detection of kidney tubular injury. *Am. J. Physiol. Renal Physiol.*, **290**, F517–F529.

39 Tonomura, Y., Tsuchiya, N., Torii, M., and Uehara, T. (2010) Evaluation of the usefulness of urinary biomarkers for nephrotoxicity in rats. *Toxicology*, **273**, 53–59.

40 Sasaki, D., Yamada, A., Umeno, H., Kurihara, H., Nakatsuji, S., Fujihira, S., Tsubota, K., Ono, M., Moriguchi, A., Watanabe, K., and Seki, J. (2011) Comparison of the course of biomarker changes and kidney injury in a rat model of drug-induced acute kidney injury. *Biomarkers*, **16**, 553–566.

41 Vinken, P., Starckx, S., Barale-Thomas, E., Looszova, A., Sonee, M., Goeminne, N., Versmissen, L., Buyens, K., and Lampo, A. (2012) Tissue Kim-1 and urinary clusterin as early indicators of cisplatin-induced acute kidney injury in rats. *Toxicol. Pathol.*, **40**, 1049–1062.

42 Pinches, M., Betts, C., Bickerton, S., Burdett, L., Thomas, H., Derbyshire, N., Jones, H.B., and Moores, M. (2012) Evaluation of novel renal biomarkers with a cisplatin model of kidney injury: gender and dosage differences. *Toxicol. Pathol.*, **40**, 522–533.

43 Fuchs, T.C., Frick, K., Emde, B., Czasch, S., von Landenberg, F., and Hewitt, P. (2012) Evaluation of novel acute urinary rat kidney toxicity biomarker for subacute toxicity studies in preclinical trials. *Toxicol. Pathol.*, **40**, 1031–1048.

44 Wadey, R.M., Pinches, M.G., Jones, H.B., Riccardi, D., and Price, S.A. (2014) Tissue expression and correlation of a panel of urinary biomarkers following cisplatin-induced kidney injury. *Toxicol. Pathol.*, **42**, 591–602.

45 McDuffie, J.E., Ma, J.Y., Sablad, M., Sonee, M., Varacallo, L., Louden, C., Guy, A., Vegas, J., Liu, X., La, D., and Snook, S. (2013) Time course of renal proximal tubule injury, reversal, and related biomarker changes in rats following cisplatin administration. *Int. J. Toxicol.*, **32**, 251–260.

46 Pavkovic, M., Riefke, B., Gutberlet, K., Raschke, M., and Ellinger-Ziegelbauer, H. (2014) Comparison of the MesoScale Discovery and Luminex multiplex platforms for measurement of urinary biomarkers in a cisplatin rat kidney injury model. *J. Pharmacol. Toxicol. Methods*, **69**, 196–204.

47 Sinha, V., Vence, L.M., and Salahudeen, A.K. (2013) Urinary tubular protein-based biomarkers in the rodent model of cisplatin nephrotoxicity: a comparative analysis of serum creatinine, renal histology, and urinary KIM-1, NGAL, and NAG in the initiation, maintenance, and recovery phases of acute kidney injury. *J. Investig. Med.*, **61**, 564–568.

48 Zhou, Y., Vaidya, V.S., Brown, R.P., Zhang, J., Rosenzweig, B.A., Thompson, K.L., Miller, T.J., Bonventre, J.V., and Goering, P.L. (2008) Comparison of kidney injury molecule-1 and other nephrotoxicity biomarkers in urine and kidney following acute exposure to gentamicin, mercury, and chromium. *Toxicol. Sci.*, **101**, 159–170.

49 Sieber, M., Hoffmann, D., Adler, M., Vaidya, V.S., Clement, M., Bonventre, J.V., Zidek, N., Rached, E., Amberg, A., Callanan, J.J., Dekant, W., and Mally, A. (2009) Comparative analysis of novel noninvasive renal biomarkers and metabonomic changes in a rat model of gentamicin nephrotoxicity. *Toxicol. Sci.*, **109**, 336–349.

50 Hoffmann, D., Adler, M., Vaidya, V., Rached, E., Mulrane, L., Gallagher, W.M., Callanan, J.J., Gautier, J.C., Matheis, K., Staedtler, F., Dieterle, F., Brandenburg, A., Sposny, A., Hewitt, P., Ellinger-Ziegelbauer, H., Bonventre, J.V., Dekant, W., and Mally, A. (2010) Performance of novel kidney biomarkers in preclinical toxicity studies. *Toxicol. Sci.*, **116**, 8–22.

51 Rached, E., Hoffmann, D., Blumbach, K., Weber, K., Dekant, W., and Mally, A. (2008) Evaluation of putative biomarkers of nephrotoxicity after exposure to ochratoxin A *in vivo* and *in vitro*. *Toxicol. Sci.*, **103**, 371–381.

52 Bandele, O., Camacho, L., Ferguson, M., Reimschuessel, R., Stine, C., Black, T., Olejnik, N., Keltner, Z., Scott, M., Gamboa da Costa, G., and Sprando, R. (2013) Performance of urinary and gene expression biomarkers in detecting the nephrotoxic effects of melamine and cyanuric acid following diverse scenarios of co-exposure. *Food Chem. Toxicol.*, **51**, 106–113.

53 Prozialeck, W.C., Edwards, J.R., Vaidya, V.S., and Bonventre, J.V. (2009) Preclinical evaluation of novel urinary biomarkers of cadmium nephrotoxicity. *Toxicol. Appl. Pharmacol.*, **238**, 301–305.

54 Wunnapuk, K., Liu, X., Peake, P., Gobe, G., Endre, Z., Grice, J.E., Roberts, M.S., and Buckley, N.A. (2013) Renal biomarkers predict nephrotoxicity after paraquat. *Toxicol. Lett.*, **222**, 280–288.

55 Wunnapuk, K., Gobe, G., Endre, Z., Peake, P., Grice, J.E., Roberts, M.S., Buckley, N.A., and Liu, X. (2014) Use of a glyphosate-based herbicide-induced nephrotoxicity model to investigate a panel of kidney injury biomarkers. *Toxicol. Lett.*, **225**, 192–200.

56 Burt, D., Crowell, S.J., Ackley, D.C., Magee, T.V., and Aubrecht, J. (2014) Application of emerging biomarkers of acute kidney injury in development of kidney-sparing polypeptide-based antibiotics. *Drug Chem. Toxicol.*, **37**, 204–212.

57 Gautier, J.C., Riefke, B., Walter, J., Kurth, P., Mylecraine, L., Guilpin, V., Barlow, N., Gury, T., Hoffman, D., Ennulat, D., Schuster, K., Harpur, E., and Pettit, S. (2010) Evaluation of novel biomarkers of nephrotoxicity in two strains of rat treated with cisplatin. *Toxicol. Pathol.*, **38**, 943–956.

58 Harpur, E., Ennulat, D., Hoffman, D., Betton, G., Gautier, J.C., Riefke, B., Bounous, D., Schuster, K., Beushausen, S., Guffroy, M., Shaw, M., Lock, E., and Pettit, S. (2011) Biological qualification of biomarkers of chemical-induced renal toxicity in two strains of male rat. *Toxicol. Sci.*, **122**, 235–252.

59 Dieterle, F., Perentes, E., Cordier, A., Roth, D.R., Verdes, P., Grenet, O., Pantano, S., Moulin, P., Wahl, D., Mahl, A., End, P., Staedtler, F., Legay, F., Carl, K., Laurie, D., Chibout, S.D., Vonderscher, J., and Maurer, G. (2010) Urinary clusterin, cystatin C, beta2-microglobulin and total protein as markers to detect drug-induced kidney injury. *Nat. Biotechnol.*, **28**, 463–469.

60 Dodiya, H., Jain, M., and Goswami, S.S. (2011) Renal toxicity of lisinopril and rosuvastatin, alone and in combination, in Wistar rats. *Int. J. Toxicol.*, **30**, 518–527.

61 Kai, K., Yamaguchi, T., Yoshimatsu, Y., Kinoshita, J., Teranishi, M., and Takasaki, W. (2013) Neutrophil gelatinase-associated lipocalin, a sensitive urinary biomarker of acute kidney injury in dogs receiving gentamicin. *J. Toxicol. Sci.*, **38**, 269–277.

62 Garcia-Martinez, J.D., Tvarijonaviciute, A., Ceron, J.J., Caldin, M., and Martinez-Subiela, S. (2012) Urinary clusterin as a renal marker in dogs. *J. Vet. Diagn. Invest.*, **24**, 301–306.

63 Keirstead, N.D., Wagoner, M.P., Bentley, P., Blais, M., Brown, C., Cheatham, L., Ciaccio, P., Dragan, Y., Ferguson, D., Fikes, J., Galvin, M., Gupta, A., Hale, M., Johnson, N., Luo, W., McGrath, F.,

Pietras, M., Price, S., Sathe, A.G., Sasaki, J.C., Snow, D., Walsky, R.L., and Kern, G. (2014) Early prediction of polymyxin-induced nephrotoxicity with next-generation urinary kidney injury biomarkers. *Toxicol. Sci.*, **137**, 278–291.

64 Vlasakova, K., Erdos, Z., Troth, S.P., McNulty, K., Chapeau-Campredon, V., Mokrzycki, N., Muniappa, N., Gu, Y.Z., Holder, D., Bailey, W.J., Sistare, F.D., and Glaab, W.E. (2014) Evaluation of the relative performance of 12 urinary biomarkers for renal safety across 22 rat sensitivity and specificity studies. *Toxicol. Sci.*, **138**, 3–20.

65 Vaidya, V.S., Ozer, J.S., Dieterle, F., Collings, F.B., Ramirez, V., Troth, S., Muniappa, N., Thudium, D., Gerhold, D., Holder, D.J., Bobadilla, N.A., Marrer, E., Perentes, E., Cordier, A., Vonderscher, J., Maurer, G., Goering, P.L., Sistare, F.D., and Bonventre, J.V. (2010) Kidney injury molecule-1 outperforms traditional biomarkers of kidney injury in preclinical biomarker qualification studies. *Nat. Biotechnol.*, **28**, 478–485.

66 Endre, Z.H., Pickering, J.W., Walker, R.J., Devarajan, P., Edelstein, C.L., Bonventre, J.V., Frampton, C.M., Bennett, M.R., Ma, Q., Sabbisetti, V.S., Vaidya, V.S., Walcher, A.M., Shaw, G.M., Henderson, S.J., Nejat, M., Schollum, J.B., and George, P.M. (2011) Improved performance of urinary biomarkers of acute kidney injury in the critically ill by stratification for injury duration and baseline renal function. *Kidney Int.*, **79**, 1119–1130.

67 Fontanilla, J. and Han, W.K. (2011) Kidney injury molecule-1 as an early detection tool for acute kidney injury and other kidney diseases. *Expert Opin. Med. Diagn.*, **5**, 161–173.

68 Gil, H.W., Yang, J.O., Lee, E.Y., and Hong, S.Y. (2009) Clinical implication of urinary neutrophil gelatinase-associated lipocalin and kidney injury molecule-1 in patients with acute paraquat intoxication. *Clin. Toxicol.*, **47**, 870–875.

69 Han, W.K., Wagener, G., Zhu, Y., Wang, S., and Lee, H.T. (2009) Urinary biomarkers in the early detection of acute kidney injury after cardiac surgery. *Clin. J. Am. Soc. Nephrol.*, **4**, 873–882.

70 Liang, X.L., Liu, S.X., Chen, Y.H., Yan, L.J., Li, H., Xuan, H.J., Liang, Y.Z., and Shi, W. (2010) Combination of urinary kidney injury molecule-1 and interleukin-18 as early biomarker for the diagnosis and progressive assessment of acute kidney injury following cardiopulmonary bypass surgery: a prospective nested case-control study. *Biomarkers*, **15**, 332–339.

71 Perco, P. and Oberbauer, R. (2008) Kidney injury molecule-1 as a biomarker of acute kidney injury in renal transplant recipients. *Nat. Clin. Pract. Nephrol.*, **4**, 362–363.

72 Abassi, Z., Shalabi, A., Sohotnik, R., Nativ, O., Awad, H., Bishara, B., Frajewicki, V., Sukhotnik, I., and Abbasi, A. (2012) Urinary NGAL and KIM-1: biomarkers for assessment of acute ischemic kidney injury following nephron sparing surgery. *J. Urol.*, **189**, 1559–1566.

73 Xue, W., Xie, Y., Wang, Q., Xu, W., Mou, S., and Ni, Z. (2014) Diagnostic performance of urinary kidney injury molecule-1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL) for acute kidney injury in obstructive nephropathy patients. *Nephrology*, **19**, 186–194.

74 Mishra, J., Ma, Q., Kelly, C., Mitsnefes, M., Mori, K., Barasch, J., and Devarajan, P. (2006) Kidney NGAL is a novel early marker of acute injury following transplantation. *Pediatr. Nephrol.*, **21**, 856–863.

75 Niemann, C.U., Walia, A., Waldman, J., Davio, M., Roberts, J.P., Hirose, R., and Feiner, J. (2009) Acute kidney injury during liver transplantation as determined by neutrophil gelatinase-associated lipocalin. *Liver Transpl.*, **15**, 1852–1860.

76 Dent, C.L., Ma, Q., Dastrala, S., Bennett, M., Mitsnefes, M.M., Barasch, J., and Devarajan, P. (2007) Plasma neutrophil gelatinase-associated lipocalin predicts acute kidney injury, morbidity and mortality after pediatric cardiac surgery: a prospective uncontrolled cohort study. *Crit. Care*, **11**, R127.

77 Bennett, M., Dent, C.L., Ma, Q., Dastrala, S., Grenier, F., Workman, R., Syed, H.,

Ali, S., Barasch, J., and Devarajan, P. (2008) Urine NGAL predicts severity of acute kidney injury after cardiac surgery: a prospective study. *Clin. J. Am. Soc. Nephrol.*, **3**, 665–673.

78 Wheeler, D.S., Devarajan, P., Ma, Q., Harmon, K., Monaco, M., Cvijanovich, N., and Wong, H.R. (2008) Serum neutrophil gelatinase-associated lipocalin (NGAL) as a marker of acute kidney injury in critically ill children with septic shock. *Crit. Care Med.*, **36**, 1297–1303.

79 Krawczeski, C.D., Woo, J.G., Wang, Y., Bennett, M.R., Ma, Q., and Devarajan, P. (2011) Neutrophil gelatinase-associated lipocalin concentrations predict development of acute kidney injury in neonates and children after cardiopulmonary bypass. *J. Pediatr.*, **158**, 1009.e1–1015.e1.

80 El-Farghali, O.G., El-Raggal, N.M., Mahmoud, N.H., and Zaina, G.A. (2012) Serum neutrophil gelatinase-associated lipocalin as a predictor of acute kidney injury in critically-ill neonates. *Pak. J. Biol. Sci.*, **15**, 231–237.

81 Makris, K., Markou, N., Evodia, E., Dimopoulou, E., Drakopoulos, I., Ntetsika, K., Rizos, D., Baltopoulos, G., and Haliassos, A. (2009) Urinary neutrophil gelatinase-associated lipocalin (NGAL) as an early marker of acute kidney injury in critically ill multiple trauma patients. *Clin. Chem. Lab. Med.*, **47**, 79–82.

82 Gaspari, F., Cravedi, P., Mandala, M., Perico, N., de Leon, F.R., Stucchi, N., Ferrari, S., Labianca, R., Remuzzi, G., and Ruggenenti, P. (2010) Predicting cisplatin-induced acute kidney injury by urinary neutrophil gelatinase-associated lipocalin excretion: a pilot prospective case-control study. *Nephron Clin. Pract.*, **115**, c154–c160.

83 Amin, R.P., Vickers, A.E., Sistare, F., Thompson, K.L., Roman, R.J., Lawton, M., Kramer, J., Hamadeh, H.K., Collins, J., Grissom, S., Bennett, L., Tucker, C.J., Wild, S., Kind, C., Oreffo, V., Davis, J.W., 2nd, Curtiss, S., Naciff, J.M., Cunningham, M., Tennant, R., Stevens, J., Car, B., Bertram, T.A., and Afshari, C.A. (2004) Identification of putative gene based markers of renal toxicity. *Environ. Health Perspect.*, **112**, 465–479.

84 Thukral, S.K., Nordone, P.J., Hu, R., Sullivan, L., Galambos, E., Fitzpatrick, V.D., Healy, L., Bass, M.B., Cosenza, M.E., and Afshari, C.A. (2005) Prediction of nephrotoxicant action and identification of candidate toxicity-related biomarkers. *Toxicol. Pathol.*, **33**, 343–355.

85 Kharasch, E.D., Schroeder, J.L., Bammler, T., Beyer, R., and Srinouanprachanh, S. (2006) Gene expression profiling of nephrotoxicity from the sevoflurane degradation product fluoromethyl-2,2-difluoro-1-(trifluoromethyl)vinyl ether ("compound A") in rats. *Toxicol. Sci.*, **90**, 419–431.

86 Hoffmann, D., Adler, M., Vaidya, V., Rached, E., Mulrane, L., Gallagher, W.M., Callanan, J.J., Gautier, J.C., Matheis, K., Staedtler, F., Dieterle, F., Brandenburg, A., Sposny, A., Hewitt, P., Ellinger-Ziegelbauer, H., Bonventre, J.V., Dekant, W., and Mally, A. (2010) Performance of novel kidney biomarkers in preclinical safety studies. *Toxicol. Sci.*, **116**, 8–22.

87 Wang, E.J., Snyder, R.D., Fielden, M.R., Smith, R.J., and Gu, Y.Z. (2008) Validation of putative genomic biomarkers of nephrotoxicity in rats. *Toxicology*, **246**, 91–100.

88 Kondo, C., Minowa, Y., Uehara, T., Okuno, Y., Nakatsu, N., Ono, A., Maruyama, T., Kato, I., Yamate, J., Yamada, H., Ohno, Y., and Urushidani, T. (2009) Identification of genomic biomarkers for concurrent diagnosis of drug-induced renal tubular injury using a large-scale toxicogenomics database. *Toxicology*, **265**, 15–26.

89 Matheis, K.A., Com, E., Gautier, J.C., Guerreiro, N., Brandenburg, A., Gmuender, H., Sposny, A., Hewitt, P., Amberg, A., Boernsen, O., Riefke, B., Hoffmann, D., Mally, A., Kalkuhl, A., Suter, L., Dieterle, F., and Staedtler, F. (2011) Cross-study and cross-omics comparisons of three nephrotoxic compounds reveal mechanistic insights and new candidate biomarkers. *Toxicol. Appl. Pharmacol.*, **252**, 112–122.

90 Minowa, Y., Kondo, C., Uehara, T., Morikawa, Y., Okuno, Y., Nakatsu, N.,

Ono, A., Maruyama, T., Kato, I., Yamate, J., Yamada, H., Ohno, Y., and Urushidani, T. (2012) Toxicogenomic multigene biomarker for predicting the future onset of proximal tubular injury in rats. *Toxicology*, **297**, 47–56.

91 Li, Y.F., Jing, Y., Hao, J., Frankfort, N.C., Zhou, X., Shen, B., Liu, X., Wang, L., and Li, R. (2013) MicroRNA-21 in the pathogenesis of acute kidney injury. *Protein Cell*, **4**, 813–819.

92 Saikumar, J., Hoffmann, D., Kim, T.M., Gonzalez, V.R., Zhang, Q., Goering, P.L., Brown, R.P., Bijol, V., Park, P.J., Waikar, S.S., and Vaidya, V.S. (2012) Expression, circulation, and excretion profile of microRNA-21, -155, and -18a following acute kidney injury. *Toxicol. Sci.*, **129**, 256–267.

93 Ben-Dov, I.Z., Tan, Y.C., Morozov, P., Wilson, P.D., Rennert, H., Blumenfeld, J.D., and Tuschl, T. (2014) Urine microRNA as potential biomarkers of autosomal dominant polycystic kidney disease progression: description of miRNA profiles at baseline. *PLoS One*, **9**, e86856.

94 Zhou, H., Hasni, S.A., Perez, P., Tandon, M., Jang, S.I., Zheng, C., Kopp, J.B., Austin, H., 3rd, Balow, J.E., Alevizos, I., and Illei, G.G. (2013) miR-150 promotes renal fibrosis in lupus nephritis by downregulating SOCS1. *J. Am. Soc. Nephrol.*, **24**, 1073–1087.

95 Yang, Y., Xiao, L., Li, J., Kanwar, Y.S., Liu, F., and Sun, L. (2013) Urine miRNAs: potential biomarkers for monitoring progression of early stages of diabetic nephropathy. *Med. Hypotheses*, **81**, 274–278.

96 Silva-Santos, R.M., Costa-Pinheiro, P., Luis, A., Antunes, L., Lobo, F., Oliveira, J., Henrique, R., and Jeronimo, C. (2013) MicroRNA profile: a promising ancillary tool for accurate renal cell tumour diagnosis. *Br. J. Cancer*, **109**, 2646–2653.

97 Chen, J., Zhang, D., Zhang, W., Tang, Y., Yan, W., Guo, L., and Shen, B. (2013) Clear cell renal cell carcinoma associated microRNA expression signatures identified by an integrated bioinformatics analysis. *J. Transl. Med.*, **11**, 169.

98 Dear, J.W., Street, J.M., and Bailey, M.A. (2013) Urinary exosomes: a reservoir for biomarker discovery and potential mediators of intrarenal signalling. *Proteomics*, **13**, 1572–1580.

99 Lv, L.L., Cao, Y., Liu, D., Xu, M., Liu, H., Tang, R.N., Ma, K.L., and Liu, B.C. (2013) Isolation and quantification of microRNAs from urinary exosomes/microvesicles for biomarker discovery. *Int. J. Biol. Sci.*, **9**, 1021–1031.

100 Lindsay, M.A. (2008) microRNAs and the immune response. *Trends Immunol.*, **29**, 343–351.

101 Faraoni, I., Antonetti, F.R., Cardone, J., and Bonmassar, E. (2009) miR-155 gene: a typical multifunctional microRNA. *Biochim. Biophys. Acta*, **1792**, 497–505.

102 Wang, N., Zhou, Y., Jiang, L., Li, D., Yang, J., Zhang, C.Y., and Zen, K. (2012) Urinary microRNA-10a and microRNA-30d serve as novel, sensitive and specific biomarkers for kidney injury. *PLoS One*, **7**, e51140.

103 Dieterle, F., Sistare, F., Goodsaid, F., Papaluca, M., Ozer, J.S., Webb, C.P., Baer, W., Senagore, A., Schipper, M.J., Vonderscher, J., Sultana, S., Gerhold, D.L., Phillips, J.A., Maurer, G., Carl, K., Laurie, D., Harpur, E., Sonee, M., Ennulat, D., Holder, D., Andrews-Cleavenger, D., Gu, Y.Z., Thompson, K.L., Goering, P.L., Vidal, J.M., Abadie, E., Maciulaitis, R., Jacobson-Kram, D., Defelice, A.F., Hausner, E.A., Blank, M., Thompson, A., Harlow, P., Throckmorton, D., Xiao, S., Xu, N., Taylor, W., Vamvakas, S., Flamion, B., Lima, B.S., Kasper, P., Pasanen, M., Prasad, K., Troth, S., Bounous, D., Robinson-Gravatt, D., Betton, G., Davis, M.A., Akunda, J., McDuffie, J.E., Suter, L., Obert, L., Guffroy, M., Pinches, M., Jayadev, S., Blomme, E.A., Beushausen, S.A., Barlow, V.G., Collins, N., Waring, J., Honor, D., Snook, S., Lee, J., Rossi, P., Walker, E., and Mattes, W. (2010) Renal biomarker qualification submission: a dialog between the FDA-EMEA and Predictive Safety Testing Consortium. *Nat. Biotechnol.*, **28**, 455–462.

104 FDA (2014) Nonclinical Qualification of Urinary Biomarkers of Nephrotoxicity.

Available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DrugDevelopmentTools QualificationProgram/UCM285010.pdf (last accessed February 10, 2014).

105 Fuchs, T.C., Mally, A., Wool, A., Beiman, M., and Hewitt, P. (2014) An exploratory evaluation of the utility of transcriptional and urinary kidney injury biomarkers for the prediction of aristolochic acid-induced renal injury in male rats. *Vet. Pathol.*, **51**, 680–694.

106 John-Baptiste, A., Vitsky, A., Sace, F., Zong, Q., Ko, M., Yafawi, R., and Liu, L. (2012) Comparison of 3 kidney injury multiplex panels in rats. *Int. J. Toxicol.*, **31**, 529–536.

107 van de Vrie, M., Deegens, J.K., van der Vlag, J., and Hilbrands, L.B. (2014) Effect of long-term storage of urine samples on measurement of kidney injury molecule 1 (KIM-1) and neutrophil gelatinase-associated lipocalin (NGAL). *Am. J. Kidney Dis.*, **63**, 573–576.

108 Parikh, C.R., Butrymowicz, I., Yu, A., Chinchilli, V.M., Park, M., Hsu, C.Y., Reeves, W.B., Devarajan, P., Kimmel, P.L., Siew, E.D., and Liu, K.D. (2014) Urine stability studies for novel biomarkers of acute kidney injury. *Am. J. Kidney Dis.*, **63**, 567–572.

10
The Use of Renal Cell Culture for Nephrotoxicity Investigations

Anja Wilmes and Paul Jennings

10.1
Introduction

The kidneys are vital organs, which continuously filter plasma into their tubules and reabsorb essential substances such as sodium, glucose, and amino acids, while waste products and excess substances are eliminated in the urine. In this way, the kidney finely regulates the constituents and volume of the blood and is the main regulator of whole body homeostasis. Due to the multitude of transporting and metabolizing systems required to perform these tasks, cells of the kidney interact with a wide variety of chemicals entities. This, coupled with its ability to concentrate compounds, makes it second only to the liver as the organ most adversely affected by xenobiotics. Detection of renal injury processes is complicated by the fact that it has a very large functional reserve. Thus, renal injury is difficult to detect before this functional reserve has been breached. Unlike the liver, the kidney has limited regenerative capacity. Also, since there is no *de novo* nephrogenesis in adults, nephrons are continually lost throughout life. The incidence of chronic kidney disease (CKD) is rising due to factors including higher prevalence of diabetes and cardiovascular disease, obesity, and aging populations [1,2].

The kidney is a highly complex organ with an estimated 15 morphologically distinct cell types [3,4]. The cells of the nephron exhibit a high degree of physiological, morphological, and biochemical heterogeneity. Due to the discrete properties of the individual cells and their immediate environment, nephron segments exhibit site-specific sensitivities to xenobiotics. Of these, the glomerulus and the proximal tubule are the most frequent nephron regions injured by drugs and chemicals and thus are the most studied in the context of disease and drug safety assessment. The glomerular filtration barrier is composed of microvascular endothelial cells, the glomerular basement membrane (GBM), and visceral epithelial cells, also known as podocytes. The properties of these components produce a glomerular filtrate devoid of cells and with a minor protein content. Podocytes have a key role in maintaining the integrity of the glomerular filter. They produce and secrete the proteins of the GBM including laminin and

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

collagen IV and form a 40 nm diameter filtration barrier through their interdigitated foot processes, which is spanned by the slit diaphragm [5,6]. They also produce the integral slit diaphragm proteins nephrin and podocin. Podocytes secrete important paracrine factors, such as vascular epithelial growth factor (VEGF) and angiopoetin-1, that influence the growth and phenotype of neighboring endothelial cells by promoting angiogenesis and vascular fenestration [6]. Due to the deep involvement of podocytes in glomerular filtration, any stimulus that disturbs their function will disrupt the filtration barrier leading to proteinuria and glomerular lesions. Protein overload will also damage the cells of the proximal tubule. Glomerular disease (or glomerulonephritis (GN)) is major contributor to CKD.

The glomerular filtrate passes directly into the lumen of the proximal tubule, where water and solutes are reabsorbed isoosmotically, from the lumen to the interstitium. In addition, specific metabolites are passed from the interstitium into the lumen and excreted in the urine. Up to two-thirds of the filtered water and solutes is reabsorbed along this nephron segment [7]. This efficient reabsorption is initiated by basolateral Na^+/K^+-ATPases, which lower intracellular Na^+ concentration allowing secondary active reabsorption, together with Na^+ in the apical membrane, or in exchange for Na^+ at the basolateral membrane. Proximal tubule cells possess a high content of mitochondria to fuel this highly energetically demanding process. The proximal tubule also recovers most of the filtered protein, through receptor-mediated endocytosis [8]. The proximal tubule expresses high amounts of organic anion transporters and organic cation transporters, which have broad substrate specificity [7,9–11]. In addition, they possess several ABC transporters, peptide transporters, nucleoside transporters, and copper transporters, all of which have been implicated in specific xenobiotic uptake [12,13]. The proximal tubule also expresses a wide variety of phase I and phase II metabolizing enzymes, including cytochrome P450 enzymes, glutathione *S*-transferases, cysteine-*S*-conjugate β-lyase, γ-glutamyl transferase, sulfotransferases, UDP-glucuronosyltransferases, flavin-containing monooxygenases, NADPH-cytochrome P450 reductases, epoxide hydrolases, and carboxylesterases [14–16]. The ability of the proximal tubule to transport a wide variety of chemical entities together with its ability to metabolize them makes it a very frequent target for drug-induced injury.

While there are other renal cells that are also important players under certain disease states and are adversely affected by xenobiotics, in this chapter we will focus predominately on podocytes and proximal tubule cells.

10.2 *In Vitro* Renal Models

Historically, *in vitro* methods to investigate physiological and pathophysiological aspects of the kidney have included techniques such as the isolated perfused kidney, isolated nephron segments, and kidney slices. For a number of reasons,

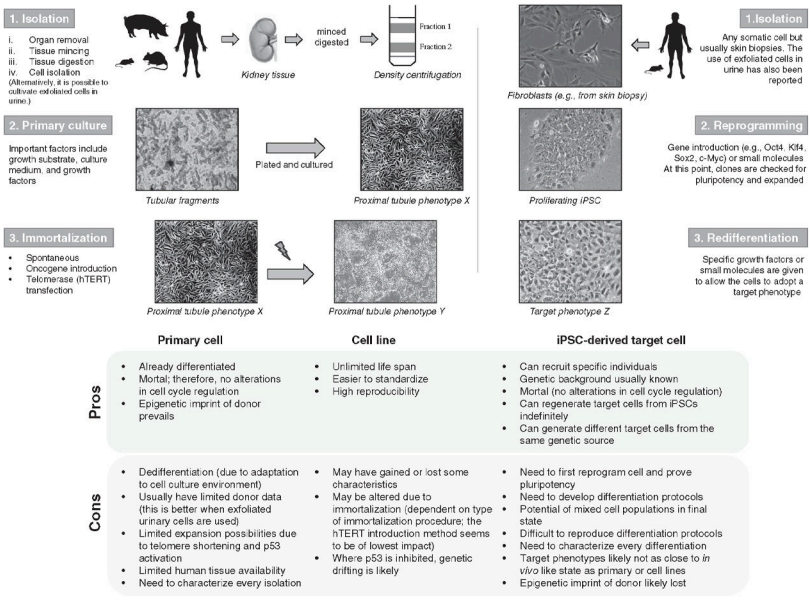
these methods have been largely replaced with cell culture techniques. Cell culture has several advantages, including the ability to conduct long-term exposures and precisely control the microenvironment. In addition, molecular biology techniques, including transcriptomic, metabolomic, and proteomic approaches, are very well suited to cell culture [17]. Furthermore, the ability to genetically modify cells in culture by introduction of modified proteins or by suppression of native genes allows the investigation of gene and protein functions under well-defined conditions. This is especially advantageous in unraveling networks and understanding mechanisms underlying the initiation and progression of disease states and adverse reactions to xenobiotics.

Renal cell culture whether it be a primary cell culture or a cell line originates with the isolation of cells from human or animal kidney tissue (Figure 10.1). The technique to isolate cortical renal cells involves removal of the renal capsule and mincing of the outer renal cortex into small pieces [18]. These pieces can be plated and cultured directly, usually leading to a high contribution of proximal tubule cells, as they are in highest abundance in the cortex [19]. This protocol has several add-ons in order to increase purity or select non-proximal tubule cells. The next step often involves enzymatic separation of the tissue pieces, for example, with collagenase, liberating glomeruli and breaking up the remaining nephron into fragments and/or cells depending on the length of digestion. Isolated glomeruli can be selected by differential sieving of the digested tissue, discarding material above ~180 μm and retaining material above ~100 μm (for humans and rats) [18]. Isolated glomeruli can then be plated onto collagen- or fibronectin-coated dishes in standard medium, for example, RPMI with FCS [20]. Glomerular microvascular endothelial cells, mesangial cells, and podocytes grow out from the glomeruli over the subsequent days; however, podocytes are usually the most numerous cell type. Where specific external antigens are known (e.g., CD31 for endothelial cells), antibodies can be used to enrich target cells or deplete unwanted cells by magnetic activated cell sorting (MACS) or fluorescent activated cell sorting (FACS).

Alternatively, or in addition to this technique, density centrifugation using, for example, Percoll can be applied to separate glomerular and tubular fragments from the collagen-digested renal material [18,21]. Washed tubular fragments can be plated on collagen- or fibronectin-coated plates in specialized medium consisting of low-glucose-containing DMEM/Ham's F-12, supplemented with epithelial growth factor (EGF), hydrocortisone, insulin, selenium, and transferrin [21,22]. Proximal tubule cells are the major outgrowths and will form a monolayer with dome-forming capacity. Domes are visualized as an out-of-focus area when viewed under a phase contrast microscope and represent an area of cells that have lifted slightly from the plate due to vectorial transport of water and solutes [22].

10.2.1 Characterization

Detailed characterization of primary cells and cell lines is essential in order to demonstrate that the correct cell type has been selected and also to determine



	Primary cell	Cell line	iPSC-derived target cell
Pros	• Already differentiated • Mortal; therefore, no alterations in cell cycle regulation • Epigenetic imprint of donor prevails	• Unlimited life span • Easier to standardize • High reproducibility	• Can recruit specific individuals • Genetic background usually known • Mortal (no alterations in cell cycle regulation) • Can regenerate target cells from iPSCs indefinitely • Can generate different target cells from the same genetic source
Cons	• Dedifferentiation (due to adaptation to cell culture environment) • Usually have limited donor data (this is better when exfoliated urinary cells are used) • Limited expansion possibilities due to telomere shortening and p53 activation • Limited human tissue availability • Need to characterize every isolation	• May have gained or lost some characteristics • May be altered due to immortalization (dependent on type of immortalization procedure; the hTERT introduction method seems to be of lowest impact) • Where p53 is inhibited, genetic drifting is likely	• Need to first reprogram cell and prove pluripotency • Need to develop differentiation protocols • Potential of mixed cell populations in final state • Difficult to reproduce differentiation protocols • Need to characterize every differentiation • Target phenotypes likely not as close to *in vivo* like state as primary or cell lines • Epigenetic imprint of donor likely lost

Figure 10.1 Schematic layout of different approaches for *in vitro* cell culture and advantages and disadvantages of each. The examples given are for renal proximal tubule cells, but could apply to any cell type. Note that depending on specific investigations, some of the advantageous and disadvantages may be reversed. See the text for details.

the level of (de)differentiation. It is unfortunately inevitable that cells will dedifferentiate and important factors to consider for improved differentiation status will be discussed later.

Podocytes have several properties that can be used to discriminate them from other glomerular cells, which may contaminate the cultures. Also, the expression of several markers is a good indication of a differentiated podocyte phenotype. The unique structures, such as the foot process and slit diaphragms, express podocyte-specific proteins and podocyte-specific protein combinations. The foot processes, which are connected via a contractile apparatus, express F-actin, myosin II, α-actinin-4, and synaptopodin [23]. The specialized slit diaphragm, which is essentially a modified adherens junction, contains zonula occludens 1 (ZO1/TJP1), P-cadherin (CDH3), CD2-assocciated protein (CD2AP) [24,25], glomerular epithelial protein 1 (GLEPP1) [26–28], nephrin [29–32], and podocin [33–35]. The presence of other proteins has been additionally used to identify podocytes in culture, including α-actinin-4 [36], ezrin [37], the transcription factor WT1 [38], and podocalyxin [39,40]. WT1 is also expressed in the developing kidney and is discussed in more detail later. Podocytes also secrete signaling molecules such as VEGF [41] and prostaglandins [42] and respond to external stimuli including angiotensin II, sodium nitroprusside, and natriuretic peptide [43,44].

Proximal tubule cells *in vivo* have a well-developed brush border, which possesses high activities of the brush border enzymes alkaline phosphatase, leucine aminopeptidase, and γ-glutamyl transferase [45,46]. While the brush border is usually not as well developed *in vitro*, cultured proximal tubule cells retain high activities of these enzymes. In addition, proximal tubule cells possess Na^+/K^+-ATPase activities, Na^+-dependent glucose, and *p*-aminohippurate transport. Proximal tubule cells are sensitive to parathyroid hormone (PTH) but not to arginine vasopressin (AVP) and this can be measured by increased cAMP levels. This property is often used to distinguish proximal from distal cells, a potential contaminating cell type [22]. Due to their role in acid–base metabolism, proximal tubule cells exhibit a pH-dependent ammoniagenesis, which can be assessed *in vitro* [22,47]. While proximal tubule cells *in vivo* perform gluconeogenesis, this capability is usually lost in cell culture [48,49]. The main function of proximal tubule cells is the constitutive transport of water and solutes; thus, they have a relatively permissive paracellular route due to the expression of the pore-forming claudins 2 and 10 [50,51]. These claudins can be measured directly at a protein level or functionally by measuring transepithelial electrical resistance (TEER). TEER measures predominantly paracellular conductance, which is typically in the region of $100\,\Omega\,cm^2$ for proximal tubule cells, but over $1000\,\Omega\,cm^2$ for more distal regions.

10.2.2
Immortalization of Primary Cells

Since primary cells have a limited life span, it is necessary to continually repeat the isolations. This raises certain problems such as tissue availability,

uncontrolled variability in human donors, and increased workload due to the need for batch characterization. Thus, most researchers favor the use of cell lines and use primary cells for verification when needed.

Normal somatic cells enter replicative senescence due to critical shortening of the telomere, the tandemly repeated hexamers at the end of mammalian chromosomes [52]. This process is governed by the p53 pathway and thus interference with telomere shortening or p53 inhibition can induce immortalization. In the intact organism, such events are associated with cancer [53]. The most widely used renal cell lines were those that arose due to unknown events during cell propagation. This is sometimes referred to as spontaneous immortalization and includes the cell lines LLC-PK_1 (Hampshire pig) [54,55], JTC-12 (cynomolgus monkey) [56,57], and OK (American opossum) [58]. These cell lines exhibit certain characteristics of the proximal tubule, but we do not know their true origins. In order to be more certain about the origin, it is possible to immortalize cells of known origin by permanently or conditionally interfering with p53 activation. This can be achieved by the introduction of viral oncogenes, including simian virus 40 (SV40) T antigen (TAg), Epstein–Barr virus (EBV), adenovirus E1A and E1B, and human papillomavirus (HPV) E6 and E7 genes. The most widely used human renal proximal tubule cell line to date, HK-2 cells, was generated by transduction of human primary proximal tubule cells with HPV E6/E7 genes [59]. The problem with this type of immortalization, however, is that p53 is required for the maintenance of DNA and chromosomal integrity and also for the regulation of energy metabolism. Thus, cell lines generated in this way are highly glycolytic, highly proliferative, and accumulate chromosomal aberrations over time. An improvement to this method is to express the viral oncogene conditionally. For example, the introduction of the tsA58 thermolabile SV40 TAg (SV40tsA58) allows the expression of the T antigen under permissive conditions at 33 °C, but it is not expressed at temperatures over 37 °C (usually 39.5 °C is recommended). Thus, cells under so-called nonpermissive conditions stop proliferating and increase their differentiation properties. However, the inhibition of p53 during proliferation and amplification will still allow chromosomal aberrations to accumulate. In addition, the culture of cells at 39 °C is far from ideal for toxicological investigations due to the overexpression of heat shock proteins. A better way to develop human cell lines is to overexpress the catalytic subunit of telomerase [22]. This enzyme maintains telomere length, so that telomeres never reach the threshold point for p53 activation. Since human somatic cells express extremely low levels of this enzyme, re-introduction effectively prevents replicative senescence without interfering with the p53 pathway. The human proximal tubule cell line, RPTEC/TERT1, was generated with hTERT introduction alone and it has been demonstrated that the cells exhibit normal activation of the p53 pathway [22,60]. In addition, RPTEC/TERT1 cells maintain a normal diploid karyotype (46, XY) with no gross chromosomal aberrations, even at high passages [22].

10.2.3 Available Podocyte and Proximal Tubule Cell Lines

A number of proximal tubule and podocyte cell lines exist from various different species. However, for the most part, human cell lines are preferred. Here follows a short description of some of the most widely used renal cell lines. All of the proximal tubule cells mentioned are available from ATCC or other commercial sources. However, until now the commercial availability of podocyte cell lines has been limited and currently none is available from ATCC. Nonetheless, a number of studies have been published utilizing animal and human podocyte cell lines. Human podocyte cell lines have been developed by transfection of primary cells with SV40TAg or SV40tsA58 alone or in combination with hTERT [61–63]. The cell line originally reported by Saleem *et al.* was generated from glomerular isolates and subsequent SV40tsA58 transfection from a nephrectomy specimen taken from a 3-year-old child [63]. They demonstrate that switching to the 37 °C nonpermissive temperature induced growth arrest and the expression of podocyte markers such as nephrin, podocin, CD2AP, and synaptopodin [63]. Sakairi *et al.* utilized exfoliated cells in human urine to generate primary isolates and subsequent cell lines, with podocyte characteristics, from two healthy volunteers and two patients with focal segmental glomerulosclerosis [62]. The cell lines were generated by cotransfection with SV40tsA58 and hTERT and all expressed synaptopodin, nestin, and CD2AP. However, podocin was absent in all cell lineages [62].

The following paragraphs describe lines that are used as proximal tubule epithelial cell models. *LLC-PK$_1$ cells* originate from a male Hampshire pig. They express activities of typical brush border membrane enzymes [64], express the enzymatic machinery for vitamin D metabolism [65,66], and exhibit Na^+-dependent apical transport of glucose, amino acids, and phosphate [67–69]. LLC-PK1 cells exhibit vectorial transport and form domes when cultured on solid supports. They exhibit a TEER of $\sim 120\,\Omega\,cm^2$ when cultured on microporous supports and have well-developed cell–cell contacts [70]. However, they do not express claudin-2, which is present in the proximal tubule *in vivo* [71,72]. They lack other typical proximal tubule characteristics such as *p*-aminohippurate transport [73] and PTH sensitivity [74].

NRK-52E cells are a clone from the so-called normal rat kidney cell line and are proposed to be of proximal tubule origin [75]. However, they lack several key proximal tubule characteristics including vectorial transport as they do not form domes. In addition, they are non-barrier forming (TEER $\sim 12\,\Omega\,cm^2$) and lack the expression of claudin-2 and several other ubiquitous claudins [76]. They are also highly glucose dependent, a characteristic not shared by proximal tubule cells *in vivo*.

HK-2 cells are a human kidney-derived cell line, which was immortalized by introducing the human papillomavirus 16 (HPV-16) E6/E7 genes [59]. HK-2 cells are positive for the brush border enzymes alkaline phosphatase, γ-glutamyl transpeptidase, and leucine aminopeptidase and they express Na^+-dependent,

phlorizin-sensitive glucose transport and responsiveness to PTH [59,77]. However, some of these characteristics may have altered over time in culture. For example, hormonal sensitivity to PTH seems to have been lost in later passages, where AVP sensitivity was acquired [22]. A big disadvantage with these cells is that they do not become contact inhibited, do not form domes, and generate only a weak transient TEER of $\sim 20\,\Omega\,cm^2$ [21,22]. In addition, they lack expression of proximal tubule typical claudins (unpublished observation).

RPTEC/TERT1 cells were one of the first epithelial cell types generated by the overexpression of catalytic unit of telomerase alone (hTERT) [22]. These cells have similar characteristics to their parent cell, including PTH responsiveness, lack of responsiveness to AVP, pH-dependent ammoniagenesis, brush border enzyme activity, and vectorial transport of water and solutes [22]. In addition, they exhibit an extremely stable expression of genes once fully contact inhibited [60]. The cells express the appropriate proximal tubule claudins, including claudins 2 and 10, and exhibit a TEER between 100 and $150\,\Omega\,cm^2$ [22,76]. In addition, once contact inhibited they exhibit a decreased glycolytic activity, an increased oxidative capacity, and reexpress several markers of proximal tubule cells *in vivo* [60].

10.3 Stem Cells

Stem cells (SCs) are capable of indefinite self-renewal and are able to give rise to progenitor cells that can differentiate into various cell lineages that lose their capacity of indefinite self-renewal. During embryogenesis, pluripotent embryonic stem cells (ESCs) can give rise to cell lineages of all three germ layers. However, even in the adult organism, several multipotent or monopotent SCs reside at various niches throughout the body [78]. These adult SCs can typically differentiate into several different lineages and supply an important pool of cells to replace senescent cells or cells that have been damaged after injury. Examples for adult SCs niches include the hippocampus area, intestinal crypts, the canals of Hering, hair bulges, skeletal muscle fibers, and the bone marrow [78]. In the mammalian kidney, no postnatal stem cells have been reported that can give rise to all cell types of the nephron and embryonic nephron progenitor cells are exhausted before birth [79]. Therefore, *de novo* nephrogenesis cannot occur after birth. However, in the event of renal injury, repair responses and replacement of damaged cells have to be initiated. There is currently a debate whether this replacement of damaged cells is dependent on or independent of a progenitor cell population. In the last decade, evidence for an adult mammalian renal progenitor cell has been reported [80–82]. It is believed that these renal progenitors localize in the Bowman's capsule, and they have been reported to express CD133, a stem cell antigen, PAX2, a transcription factor that is normally expressed during renal development, and the glycoprotein CD24 [80,82]. However, these progenitor cells show only limited self-renewal and exhibit limited

differentiation properties. On the other hand, it has been shown that surviving epithelial cells can replace damaged tubular epithelium after acute kidney injury, completely independent of any progenitor line [83,84]. Humphreys *et al.* showed that after renal injury, nearly all renal proximal tubule cells were capable of proliferation. In addition, these proliferating cells re-expressed PAX2, which has previously been shown to be re-activated upon renal injury [84]. Another potential role in renal repair is the recruitment of external SCs that reside outside the kidney, for example, mesenchymal SCs [85–87]. There are several potential ways the kidney can repair cells in the nephron after injury, but it cannot replace lost nephrons. Thus, the kidney has only limited options of self-repair mechanisms, compared with the liver that is capable of self-regeneration, highlighting the importance of early detection of renal toxicity.

The potential use of SCs, especially adult-derived SCs, as a source for the production of differentiated target cell lines is alluring. The advantages of such a possibility would be enormous. It would allow an unlimited supply of source material, would include the possibility of including genetic variability, and allow us to readily culture target cells from diseased individuals (Figure 10.1). The stem cell field was revolutionized in 2006, when Takahashi and Yamanaka showed that mouse fibroblasts could be reprogrammed into inducible pluripotent stem cells (iPSCs) by addition of the four transcription factors Oct3/4, Klf4, Sox2, and c-Myc [88]. One year later, they described the first iPSCs that had been reprogrammed from human fibroblasts [89]. Several other sources of somatic cells have been used to develop iPSCs, including exfoliated renal epithelial cells found in the urine [90]. Improvements and optimization of iPSC protocols are ongoing and include the development of nonintegrating techniques such as the use of non-nuclear Sendai virus delivery and small-molecule strategies. The major challenge now is to develop reproducible differentiation strategies to generate target cells of, among others, proximal tubule epithelial cells and podocyte phenotypes. In order to be able to do that, we need to understand the processes involved in kidney development.

In the mammalian embryo, the first step of gastrulation is the formation of the primitive streak. Growth factors proposed to induce this event include members of the transforming growth factor beta (TGFβ) families and the Wnt families. The primitive streak can develop into both mesoderm and endoderm, and in mouse ESCs, Wnt and low levels of activin induce the posterior streak (mesoderm), whereas high levels of activin induce the anterior streak (endoderm) [91]. In addition, bone morphogenetic protein 4 (BMP4) also contributes to anterior visceral endoderm development [92]. The definitive mesoderm can then develop into the paraxial mesoderm (PM), the intermediate mesoderm (IM), and the lateral plate mesoderm (LPM). The IM will eventually give rise to the urogenital system. While the exact mechanism of IM formation remains unclear, it is believed that the IM arises in response to BMPs that are secreted by the ectoderm and by activin A and other molecules secreted by the PM [93]. From the IM, three types of kidneys develop: the pronephros develop at days 21–22 postcoitum (pc) in humans and represent a transient state, the mesonephros begin to

form at day 25 pc in humans and represent the functional embryonic kidney, and finally the adult permanent form, the metanephros, begins to form at day 30 pc in humans [94]. Pronephros are the type of kidney that is present in adult lower vertebrates, including fish and amphibians. These early stages of renal development depend on retinoic acid (RA) [95]. The mesonephros already contain glomeruli and convoluted proximal tubule-like structures but are then replaced during development by the metanephros, initiated with the outgrowth of the ureteric bud into the metanephric mesenchyme (MM) [96]. During this outgrowth, branches are formed and mesenchymal cells close to the tips of these branches begin mesenchyme-to-epithelial transitions, whereas mesenchymal cells at other parts form interstitial stroma [96]. The ureteric bud develops into the collecting duct, whereas mesenchymal cells at the tips develop into renal tubules. The outgrowth of the ureteric bud is regulated by glial cell line-derived neurotrophic factor (Gdnf) as well as by other growth factors including BMP2/7 and TGFβ [97]. In addition, fibroblast growth factor 2 (FGF2) is present at the tips of the branches and has been proposed to be involved in metanephric mesenchyme development [98].

A number of genes have been identified to be activated during early kidney development. One of the earliest genes expressed is the TF odd-skipped related 1 (Odd1) that can be detected in undifferentiated kidney precursor tissue [94]. Another early gene detected is Lim1, which is present in the visceral endoderm, anterior mesendoderm, IM, and LPM [99]. Later on, the IM specific markers PAX2 and PAX8 are expressed and all of the above, Odd1, Lim1, and PAX2/8, are necessary for correct development of the kidney [96]. After that, genes that regulate the outgrowth of the ureteric bud, including the tyrosine kinase receptor RET and its ligand Gdnf, and the coreceptor GFRα-1 are expressed [100]. The HOX family, in particular HOX11, seems to play a particular important role in anterior–posterior patterning and deletion of HOX11 resulted in disturbances of Gdnf expression and ureteric bud formation [101]. Two additional genes that are necessary for MM induction are Six1 and Eya1 [102]. Other genes that are involved in regulation of ureteric bud outgrowth and highly expressed in the MM include Wilms tumor suppressor gene WT1 and Sal1 [96].

Several studies have reported the differentiation of mouse ESCs or embryoid bodies (EBs) into renal progenitor-like cell lineages [103–105]. Addition of the combination of activin A and RA to murine EBs increased PAX2 and WT1 levels more than 20-fold and at the same time the loss of the pluripotent gene Oct4 was observed [104]. In addition, RA increased the gene expression of Six2, Gdnf, and Eya1. Further treatment with BMP7 suppressed PAX2 and WT1 and increased the expression of Eya1, Gdnf, lim1, and cadherin 6. Cadherin 6 is believed to be an early marker of proximal tubule precursors [104]. Injection of these *in vitro* generated renal progenitor cells into developing kidneys *in vivo* [104] or into newborn mouse kidneys *ex vivo* [105] showed their ability to generate renal proximal tubule structures. Recent approaches have built on this approach [106,107], using a stepwise differentiation with specific subsets of

growth factors. Takasato *et al.* initially induced posterior primitive streak formation by addition of high BMP4/low activin A for 2–3 days [106]. In the second step, IM was induced from the posterior primitive streak by addition of FGF2 or FGF9 for 4 days. More than 80% of cells expressed the IM markers LHX1 and PAX2. In the final step, the combination of FGF9, BMP7, and RA was applied for further 11 days to induce MM that expressed Six2, WT1, Gdnf, and HOXD11. After 18 days, cells were resuspended into single cells and cultured to allow self-aggregation. After another 4 days of culture, differentiation into podocytes, expressing synaptopodin, and proximal tubule cells, expressing aquaporin 1 (AQP1), could be observed [106]. Xia *et al.* used a two-step protocol to commit pluripotent SCs into ureteric bud progenitor-like cell lines. Initial exposure of BMP4 and FGF2 for 2 days to induce mesoderm-committed cells was followed by RA, activin A, and BMP2 exposure for another 2 days to generate IM. After 4 days, these cells expressed OSR1 (also known as Odd1), LHX1, PAX2, and GATA3 [107].

While these studies mainly focused on generating renal progenitor-like cells, another promising approach is to direct differentiation of pluripotent SCs to podocytes or proximal tubule cells directly *in vitro*. Recently, Narayanan *et al.* reported differentiation of human ESCs into renal proximal tubule-like cells [108]. They grew ESCs on Matrigel^{TM} in a renal epithelial growth medium supplemented with FBS, EGF, hydrocortisone, insulin, triiodothyronine, and transferrin for 20 days [108]. Differentiation of ESCs into proximal tubule cells was initiated by addition of BMP2, BMP7, activin A, and RA. Upregulation of PAX2 and AQP1 was observed after 10 days. Interestingly, the combination of BMP2 and BMP7 alone was sufficient to differentiate the cells into proximal tubule-like cells. However, the highest yield of AQP1-positive cells (~38%) was achieved when all four factors were combined. To characterize the AQP1-positive cells, several proximal tubule markers were analyzed, including kidney-specific cadherin, megalin, glutamyl transferase (GGT), aminopeptidase N (CD13), and several proximal tubule-specific transporters. In addition, functional characterization was carried out by measuring water transport, ammonia production, and response to parathyroid hormone [108].

Song *et al.* demonstrated the possibility of human iPSC differentiation into podocyte-like cells [109]. Differentiation of iPSCs was induced by growing them in DMEM/Ham's F-12 supplemented with FBS, nonessential amino acids, β-mercaptoethanol, activin A, BMP7, and RA for 10 days. iPSC-derived podocyte-like cells showed upregulation of the podocyte-specific markers podocin, nephrin, and WT1 and downregulation of Oct3/Oct4. Functional characterization was analyzed by studying uptake of albumin and contractile response to angiotensin II [109].

More work will be required to optimize these approaches for podocyte and proximal tubule differentiation from iPSCs. Until now no method has been published to differentiate iPSCs into a proximal tubule phenotype. It is likely that the ESC method developed could also be applied to iPSCs, but this has to be investigated and adopted appropriately. In addition, protocols will have to be optimized

to increase yield, purity, and obviously differentiation status. Finally and probably most importantly, the protocols must be robust and reproducible if they are to be applicable to drug safety assessment.

10.4 Optimal Cell Culture Conditions

If we are to be successful in applying SCs to drug safety regimes, we should not ignore the enormous amount of work that has been conducted using traditional renal primary cell cultures and renal cell lines. There are certain factors that we know to be critical for differentiation status, but they need to be implemented and optimized.

Lessons have already been learned from cell culture medium development and SC researches were early adopters of advanced cell culture medium. Epithelial cell cultures were originally developed for the propagation of viruses for vaccine development and thus medium was designed primarily for the selection of proliferating cells without any regard to differentiation. These cells were (and many still are) typically grown in a base medium such as DMEM or RPMI with 10% fetal calf serum (FCS) and antibiotics. However, it has been known for some time that FCS addition is not optimal for maintenance of differentiation. In addition, FCS is not standardizable and can be problematic for certain downstream applications due to its animal origin. Thus, efforts to develop kidney cells under more defined serum-free medium began decades ago. DMEM/Ham's F-12 is one of the more complete base media, containing amino acids, vitamins, fatty acids, and nutrient sources and thus is the most common medium used for serum-free application. Studies conducted in the developing cortex extracted from human fetal kidneys have shown that EGF, insulin, and transferrin have a synergistic role in cell proliferation [110] and that hydrocortisone also acts as a mitogen [111]. Other studies have demonstrated that transferrin and prostaglandin E1 were necessary for optimal growth and that prostaglandin E1 was necessary for maximal metanephric differentiation [112]. Prostaglandin E1/E2 has also been demonstrated to increase brush border activity of cultured human proximal tubule cells [113]. A medium consisting of DMEM/Ham's F-12 with insulin, transferrin, selenium (ITS), hydrocortisone, triiodothyronine, and EGF could support proliferation of rat proximal tubule cells [114]. A number of human proximal tubule cells have been developed under similar serum-free conditions, including the HK-2 and RPTEC/TERT1 cells [22,59]. In our laboratory, we tend to favor DMEM/Ham's F-12 supplemented with ITS, EGF, and hydrocortisone for growth of both primary human proximal tubule cells and human proximal tubule cell lines [21,60,115]. Serum is also avoided in stem cell culture due to the amount of uncontrolled growth factors and cytokines. DMEM/Ham's F-12 with defined growth factors such as bFGF and TGFβ1 is also now the standard for proliferation of pluripotent stem cells and is the constituent of commercial SC media such as mTeSR [116].

Another important factor in cell culture is oxygen tension. There is the frequent misconception that cells in culture are exposed to too high oxygen as for practical reasons standard cell culture incubators only control carbon dioxide tension. Thus, there is 21% oxygen in the cell culture headspace. However, the diffusion rate of oxygen through the medium to the oxygen-consuming cells is often neglected. For adherent cells, this can be a limiting factor and many adherent cell cultures may experience actual oxygen tensions in the hypoxia range. This is due to a higher oxygen consumption rate of cells at the bottom of the Petri dish than the oxygen diffusion rate through the liquid column [117]. The extent of hypoxia will depend on the cell type and cell density. Renal proximal tubule epithelial cells are particularly responsive to alterations in oxygen tensions. It has been demonstrated that alterations in medium volume, which effectively alters oxygen delivery rate, can also markedly affect cell metabolism [118]. In addition, oxygen tension is an important regulator of glomerulogenesis and nephrogenesis in the developing kidney, through HIF-1α and HIF-2α activation [119]. Thus, it is likely that hypoxia may play a critical role in temporal development of specific cell lineages.

The kidney is a highly perfused organ and both podocytes and proximal tubule cells experience permanent perfusion conditions. These perfusion conditions are important for cell and tissue functions and are also important for how the cells are exposed to xenobiotics. Under conventional culture, there is no perfusion as medium is simply replenished every 2–3 days. Medium composition is continuously modified by the cells and thus toward the end of the feeding cycle is very different from that of fresh medium. These conditions are not conducive to a stable phenotype and time of replenishment of medium can cause huge alterations in cell gene regulation [115]. Continuous perfusion of medium solves some of these issues [120,121], although this inevitably increases experimental complexity and thus decreases assay throughput considerably. Medium perfusion with improved oxygenation can improve proximal tubule phenotype. We have shown, for example, that LLC-PK1 cells cultured in an oxygenated perfusion system increased oxidative metabolism and mitochondrial volume [120]. In recent years, there has also been an increase in the development of microfluidic systems that allow greater control of medium flow and have the possibility to use online sensors [122]. There are a number of advantages of microfluidic systems, including the possibility to interconnect different compartments with different cell types. Also since flow itself is an inducer of differentiation of endothelial and epithelial cell types, such devices may greatly improve differentiation of primary cells, cell lines, and potentially also iPSC-derived cells. It has recently been demonstrated that primary human proximal tubule cell culture in a microfluidic device exhibited enhanced polarization and primary cilia formation, and increased albumin and glucose transport [123]. In addition, this study demonstrated the applicability of the microfluidic device to nephrotoxicity testing using cisplatin as a model compound.

There has been a lot of interest and discussion regarding the topic of 2D versus 3D cell culture. Some cell types, such as hepatocytes, do much better in 3D

cultures as it reflects better their structural organization in the liver *in vivo*. For example, HepaRG cells cultured in hanging drops form 3D aggregates and showed increased production rates of albumin and glucose, as well as CYP2E1 activity, compared with 2D cultures [124]. Excellent results have also been achieved for neuronal cells by cultivating cells in aggregate form, allowing a more natural brain architecture with several distinct cell phenotypes and thus reaching a high level of structural and functional maturity [125,126]. Both of these systems have been shown to be suitable for toxicity testing studies. Aggregation, however, is not well suited to renal epithelial cells, as renal cells are exposed to apical and basolateral fluids *in vivo* and culture in aggregate form negates this possibility. However, it is possible to induce tubule formation of renal epithelial cells *in vitro*. A recent study has shown improved differentiation and increased sensitivities to nephrotoxins when human proximal tubule cells were cultured as 3D tubules [127]. This was achieved by seeding the cells in collagen/Matrigel gels [127]. A major disadvantage of this approach, however, is that there is no way currently to perfuse the tubules and thus compounds (and nutrients) can only be basolaterally applied. Also, this method is not well suited for transport studies and it is questionable whether the lumens are fluid filled at all. A tried and tested way to overcome many of these issues is the culture of renal cells on microporous supports. This has a major advantage in that nutrients are supplied from both the apical and basolateral sides. For toxicological studies, this is also important as compounds can also be applied in both compartments. Culturing cells in this way can also increase polarization and differentiation [128,129]. In addition, the growth of cells on filters allows the measurement of barrier function via TEER, which is an excellent parameter for assessing barrier integrity. Using microporous supports also allows different coculture options. For example, we have shown an increased complexity of interaction when proximal tubule cells were cultured on one side of a filter and microvascular endothelial cells on the other [21,130].

10.5 *In Vitro* Nephrotoxicity Assessment

The use of kidney cells in *in vitro* studies has been successfully utilized to investigate many clinically important nephrotoxins [12], as well as nephrotoxic compounds found in the environment and in food [131,132]. As with other tissues, classical approaches for *in vitro* nephrotoxicity have relied mostly on the quantifying viability or cytotoxicity parameters, including loss of cytosolic contents (e.g., lactate dehydrogenase), dye exclusion (e.g., trypan blue), or certain metabolic activities (e.g., MTT, resazurin). In addition, several sensitive assays for quantifying apoptosis are now routinely employed, including measurement of caspase 3 activity. Assays tailored to nephron-specific events include the measurement of proximal tubule-specific brush border enzymes in the supernatant medium and functional parameters such as TEER. TEER is in particular a very

sensitive endpoint of toxicity as only a small amount of cell death can lead to a collapse in barrier function. However, none of these assays provide deep mechanistic insights into drug-induced injury, since many different cellular events can result in apoptosis or necrosis. For this reason, there has been a shift away from the exclusive use of cytotoxicity assays, in order to focus on earlier events, which are likely to provide deeper mechanistic information. For example, using molecular biological tools, we have demonstrated that cyclosporin A (CsA)-induced ROS production results in senescence of cultured human renal proximal tubule cells, via activation of the p53 pathway [133]. Furthermore, cells can sense and react to molecular perturbations primarily through the activation of stress response pathways. A large number of these stress response pathways are activated through transcription factor master regulators, including p53, Nrf2, Hif-1α, ATF4, MTF1, and XBP1 [134]. The activation status of these pathways can be measured utilizing transcriptomics, real-time PCR, or reporter assays and provides extremely useful mechanistic information, which can aid in chemical classification and identification of molecular initiating events. For example, the Nrf2 pathway is activated in renal proximal tubule cells under ischemic conditions or when exposed to several nephrotoxins [17,135,136]. The increased expression of hemeoxygenase-1 (HO-1), NQO1, thioredoxin reductase 1 (TXNRD1), and glutamate-cysteine ligase, modifier subunit (GCLM) can be used as reporters of Nrf2 activation [17,135,136].

We have recently investigated a multi-omics approach based on transcriptomics, proteomics, and metabolomics, in an attempt to elucidate the cellular perturbations induced by CsA [17]. In long-term exposures of differentiated microporous cultured RPTEC/TERT1 cells, we could demonstrate that CsA induces both oxidative stress and ER stress, potentially due to mitochondrial disturbances. Utilizing cyclophilin B secretion and pharmaco-kinetics of CsA in the *in vitro* model, we could also demonstrate that the stress effects were not related to primary pharmacology [17]. This study demonstrates the suitability of renal cell cultures to drug safety assessment, especially since primary pharmacology could be separated from toxicological events and tissue concentrations could be calculated.

10.6 Outlook

Renal cell cultures are promising tools for nephrotoxicity studies and drug safety evaluation. The ability to create human-derived cell lines by hTERT overexpression, and thus without introducing viral oncogenes, is a major development as it allows the investigation of the effects of compounds on contact-inhibited differentiated cells [17,60]. This combined with the advances in high-content omics approaches will allow us to develop better, more sensitive systems for drug safety evaluation [17]. By conducting integrated transcriptomic, proteomic, and metabolomic investigations into cellular perturbations, we will be able to (a)

understand better how cells react to specific noxious stimuli and (b) develop mechanistic markers for these pathways [17]. It would be advantageous if some of those markers were released from cells and could also be measured in urine. Such translational markers could bridge the gap between preclinical and clinical testing. Engineering developments such as those used to create microfluidic devices will likely allow more organotypic cultures and organ–organ interactions. One can envisage a system with a liver, a renal, and a cardiac module, all interconnected with microfluidic circulation. New biological developments such as iPSCs allow the possibility of creating renal cells from specific individuals and thus allow us to study genetic variability and individual susceptibilities to compounds. This technology may even usher in an era of personalized toxicology, which could entirely change how pharmaceuticals are developed and tested.

References

1 Nugent, R.A. *et al.* (2011) The burden of chronic kidney disease on developing nations: a 21st century challenge in global health. *Nephron Clin. Pract.*, **118** (3), c269–c277.

2 Meguid El Nahas, A. and Bello, A.K. (2005) Chronic kidney disease: the global challenge. *Lancet*, **365** (9456), 331–340.

3 Anonymous (1988) A standard nomenclature for structures of the kidney. The Renal Commission of the International Union of Physiological Sciences (IUPS). *Pflugers Arch.*, **411** (1), 113–120.

4 Kriz, W. and Kaissling, B. (2000) Structural organization of the mammalian kidney, in *The Kidney: Physiology and Pathophysiology*, vol. 1, 3rd edn (eds D. Seldin and G. Giebisch), Lippincott Williams & Wilkins, Philadelphia, PA.

5 Byron, A. *et al.* (2014) Glomerular cell cross-talk influences composition and assembly of extracellular matrix. *J. Am. Soc. Nephrol.*, **25** (5), 953–966.

6 Brinkkoetter, P.T., Ising, C., and Benzing, T. (2013) The role of the podocyte in albumin filtration. *Nat. Rev. Nephrol.*, **9** (6), 328–336.

7 Lash, L.H., Putt, D.A., and Cai, H. (2006) Membrane transport function in primary cultures of human proximal tubular cells. *Toxicology*, **228** (2–3), 200–218.

8 Amsellem, S. *et al.* (2010) Cubilin is essential for albumin reabsorption in the renal proximal tubule. *J. Am. Soc. Nephrol.*, **21** (11), 1859–1867.

9 Perri, D. *et al.* (2003) The kidney – the body's playground for drugs: an overview of renal drug handling with selected clinical correlates. *Can. J. Clin. Pharmacol.*, **10** (1), 17–23.

10 Robertson, E.E. and Rankin, G.O. (2006) Human renal organic anion transporters: characteristics and contributions to drug and drug metabolite excretion. *Pharmacol. Ther.*, **109** (3), 399–412.

11 Wright, S.H. and Dantzler, W.H. (2004) Molecular and cellular physiology of renal organic cation and anion transport. *Physiol. Rev.*, **84** (3), 987–1049.

12 Jennings, P. *et al.* (2008) Renal cell culture models: contribution to the understanding of nephrotoxic mechanisms, in *Clinical Nephrotoxins* (eds M.E. De Broe, G.A. Porter, W.M. Bennett, and G. Deray), Springer, New York, pp. 223–249.

13 De Broe, M.E. and Roch-Ramel, F. (2008) Renal handling of drugs and xenobiotics, in *Clinical Nephrotoxins* (eds M.E. De Broe, G.A. Porter, W.M. Bennett, and G. Deray), Springer, New York, pp. 44–71.

14 Lohr, J.W., Willsky, G.R., and Acara, M.A. (1998) Renal drug metabolism. *Pharmacol. Rev.*, **50** (1), 107–141.

15 Lock, E.A. and Reed, C.J. (1998) Xenobiotic metabolizing enzymes of the kidney. *Toxicol. Pathol.*, **26** (1), 18–25.

16 Knights, K.M., Rowland, A., and Miners, J.O. (2013) Renal drug metabolism in humans: the potential for drug–endobiotic interactions involving cytochrome P450 (CYP) and UDP-glucuronosyltransferase (UGT). *Br. J. Clin. Pharmacol.*, **76** (4), 587–602.

17 Wilmes, A. *et al.* (2013) Application of integrated transcriptomic, proteomic and metabolomic profiling for the delineation of mechanisms of drug induced cell stress. *J. Proteomics*, **79**, 180–194.

18 Jennings, P. (2001) Novel approaches for *in vitro* nephrotoxicity testing. PhD thesis, University College Dublin.

19 Detrisac, C.J. *et al.* (1984) Tissue culture of human kidney epithelial cells of proximal tubule origin. *Kidney Int.*, **25** (2), 383–390.

20 Ni, L., Saleem, M., and Mathieson, P.W. (2012) Podocyte culture: tricks of the trade. *Nephrology*, **17** (6), 525–531.

21 Aydin, S. *et al.* (2008) Influence of microvascular endothelial cells on transcriptional regulation of proximal tubular epithelial cells. *Am. J. Physiol. Cell Physiol.*, **294** (2), C543–C554.

22 Wieser, M. *et al.* (2008) hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. *Am. J. Physiol. Renal Physiol.*, **295** (5), F1365–F1375.

23 Faul, C. *et al.* (2007) Actin up: regulation of podocyte structure and function by components of the actin cytoskeleton. *Trends Cell Biol.*, **17** (9), 428–437.

24 Li, C. *et al.* (2000) CD2AP is expressed with nephrin in developing podocytes and is found widely in mature kidney and elsewhere. *Am. J. Physiol. Renal Physiol.*, **279** (4), F785–F792.

25 Shih, N.Y. *et al.* (2001) CD2AP localizes to the slit diaphragm and binds to nephrin via a novel C-terminal domain. *Am. J. Pathol.*, **159** (6), 2303–2308.

26 Thomas, P.E. *et al.* (1994) GLEPP1, a renal glomerular epithelial cell (podocyte) membrane protein-tyrosine phosphatase. Identification, molecular cloning, and characterization in rabbit. *J. Biol. Chem.*, **269** (31), 19953–19962.

27 Wang, R. *et al.* (2000) Molecular cloning, expression, and distribution of glomerular epithelial protein 1 in developing mouse kidney. *Kidney Int.*, **57** (5), 1847–1859.

28 Wiggins, R.C. *et al.* (1995) Molecular cloning of cDNAs encoding human GLEPP1, a membrane protein tyrosine phosphatase: characterization of the GLEPP1 protein distribution in human kidney and assignment of the GLEPP1 gene to human chromosome 12p12–p13. *Genomics*, **27** (1), 174–181.

29 Holthofer, H. *et al.* (1999) Nephrin localizes at the podocyte filtration slit area and is characteristically spliced in the human kidney. *Am. J. Pathol.*, **155** (5), 1681–1687.

30 Holzman, L.B. *et al.* (1999) Nephrin localizes to the slit pore of the glomerular epithelial cell. *Kidney Int.*, **56** (4), 1481–1491.

31 Kestila, M. *et al.* (1998) Positionally cloned gene for a novel glomerular protein – nephrin – is mutated in congenital nephrotic syndrome. *Mol. Cell*, **1** (4), 575–582.

32 Ruotsalainen, V. *et al.* (1999) Nephrin is specifically located at the slit diaphragm of glomerular podocytes. *Proc. Natl. Acad. Sci. USA*, **96** (14), 7962–7967.

33 Boute, N. *et al.* (2000) NPHS2, encoding the glomerular protein podocin, is mutated in autosomal recessive steroid-resistant nephrotic syndrome. *Nat. Genet.*, **24** (4), 349–354.

34 Roselli, S. *et al.* (2002) Podocin localizes in the kidney to the slit diaphragm area. *Am. J. Pathol.*, **160** (1), 131–139.

35 Schwarz, K. *et al.* (2001) Podocin, a raft-associated component of the glomerular slit diaphragm, interacts with CD2AP and nephrin. *J. Clin. Invest.*, **108** (11), 1621–1629.

36 Kaplan, J.M. *et al.* (2000) Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nat. Genet.*, **24** (3), 251–256.

37 Hugo, C. *et al.* (1998) The plasma membrane–actin linking protein, ezrin, is

a glomerular epithelial cell marker in glomerulogenesis, in the adult kidney and in glomerular injury. *Kidney Int.*, **54** (6), 1934–1944.

38 Mundlos, S. *et al.* (1993) Nuclear localization of the protein encoded by the Wilms' tumor gene WT1 in embryonic and adult tissues. *Development*, **119** (4), 1329–1341.

39 Horvat, R. *et al.* (1986) Endothelial cell membranes contain podocalyxin – the major sialoprotein of visceral glomerular epithelial cells. *J. Cell Biol.*, **102** (2), 484–491.

40 Kerjaschki, D., Sharkey, D.J., and Farquhar, M.G. (1984) Identification and characterization of podocalyxin – the major sialoprotein of the renal glomerular epithelial cell. *J. Cell Biol.*, **98** (4), 1591–1596.

41 Kretzler, M. *et al.* (1998) Detection of multiple vascular endothelial growth factor splice isoforms in single glomerular podocytes. *Kidney Int. Suppl.*, **67**, S159–S161.

42 Kreisberg, J.I., Karnovsky, M.J., and Levine, L. (1982) Prostaglandin production by homogeneous cultures of rat glomerular epithelial and mesangial cells. *Kidney Int.*, **22** (4), 355–359.

43 Durvasula, R.V. *et al.* (2004) Activation of a local tissue angiotensin system in podocytes by mechanical strain. *Kidney Int.*, **65** (1), 30–39.

44 Sharma, R. *et al.* (1992) Vasoactive substances induce cytoskeletal changes in cultured rat glomerular epithelial cells. *J. Am. Soc. Nephrol.*, **3** (5), 1131–1138.

45 Guder, W.G., Wagner, S., and Wirthensohn, G. (1986) Metabolic fuels along the nephron: pathways and intracellular mechanisms of interaction. *Kidney Int.*, **29** (1), 41–45.

46 Ross, B. and Guder, W. (1982) Heterogeneity and compartmentation in the kidney, in *Metabolic Compartmentation* (ed. H. Sies), Academic Press, London, pp. 363–409.

47 Gstraunthaler, G.J. (1994) Ammoniagenesis in renal cell culture. A comparative study on ammonia metabolism of renal epithelial cell lines. *Contrib. Nephrol.*, **110**, 88–97.

48 Courjault-Gautier, F. *et al.* (1995) Consecutive use of hormonally defined serum-free media to establish highly differentiated human renal proximal tubule cells in primary culture. *J. Am. Soc. Nephrol.*, **5** (11), 1949–1963.

49 Toutain, H. *et al.* (1991) Biochemical, functional, and morphological characterization of a primary culture of rabbit proximal tubule cells. *Exp. Cell Res.*, **194** (1), 9–18.

50 Kirk, A. *et al.* (2010) Differential expression of claudin tight junction proteins in the human cortical nephron. *Nephrol. Dial. Transplant.*, **25** (7), 2107–2119.

51 Muto, S. *et al.* (2010) Claudin-2-deficient mice are defective in the leaky and cation-selective paracellular permeability properties of renal proximal tubules. *Proc. Natl. Acad. Sci. USA*, **107** (17), 8011–8016.

52 Moyzis, R.K. *et al.* (1988) A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA*, **85** (18), 6622–6626.

53 Liu, J. *et al.* (2013) Tumor suppressor p53 and its mutants in cancer metabolism. *Cancer Lett.* doi: 10.1016/j.canlet.2013.12.025.

54 Gstraunthaler, G., Pfaller, W., and Kotanko, P. (1985) Lack of fructose-1,6-bisphosphatase activity in LLC-PK1 cells. *Am. J. Physiol.*, **248** (1 Pt 1), C181–C183.

55 Hull, R.N., Cherry, W.R., and Weaver, G.W. (1976) The origin and characteristics of a pig kidney cell strain, LLC-PK. *In Vitro*, **12** (10), 670–677.

56 Takaoka, T. *et al.* (1962) Establishment of a cell strain, JTC-12, from cynomolgus monkey kidney tissue. *Jpn. J. Exp. Med.*, **32**, 351–368.

57 Takuwa, Y. and Ogata, E. (1985) Differentiated properties characteristic of renal proximal epithelium in a cell line derived from a normal monkey kidney (JTC-12). *In Vitro Cell Dev. Biol.*, **21** (8), 445–449.

58 Koyama, H. *et al.* (1978) Establishment and characterization of a cell line from the American opossum (*Didelphys virginiana*). *In Vitro*, **14** (3), 239–246.

59 Ryan, M.J. *et al.* (1994) HK-2: an immortalized proximal tubule epithelial cell line from normal adult human kidney. *Kidney Int.*, **45** (1), 48–57.

60 Aschauer, L. *et al.* (2013) Delineation of the key aspects in the regulation of epithelial monolayer formation. *Mol. Cell. Biol.*, **33** (13), 2535–2550.

61 Delarue, F. *et al.* (1991) Stable cell line of T-SV40 immortalized human glomerular visceral epithelial cells. *Kidney Int.*, **40** (5), 906–912.

62 Sakairi, T. *et al.* (2010) Conditionally immortalized human podocyte cell lines established from urine. *Am. J. Physiol. Renal Physiol.*, **298** (3), F557–F567.

63 Saleem, M.A. *et al.* (2002) A conditionally immortalized human podocyte cell line demonstrating nephrin and podocin expression. *J. Am. Soc. Nephrol.*, **13** (3), 630–638.

64 Netzer, A. and Gstraunthaler, G. (1993) Selective release of apical membrane enzymes from cultured renal epithelia by phosphatidylinositol-specific phospholipase C. *Renal Physiol. Biochem.*, **16** (6), 299–310.

65 Costa, E.M., Hirst, M.A., and Feldman, D. (1985) Regulation of 1,25-dihydroxyvitamin D3 receptors by vitamin D analogs in cultured mammalian cells. *Endocrinology*, **117** (5), 2203–2210.

66 Kelly, M.A. *et al.* (1985) A variant form of the 1,25-dihydroxyvitamin D3 receptor with low apparent hormone affinity in cultured monkey kidney cells (LLC-MK2). A model for tissue resistance to vitamin D. *J. Biol. Chem.*, **260** (3), 1545–1549.

67 Gstraunthaler, G. and Handler, J.S. (1987) Isolation, growth, and characterization of a gluconeogenic strain of renal cells. *Am. J. Physiol.*, **252** (2 Pt 1), C232–C238.

68 Pfaller, W., Gstraunthaler, G., and Loidl, P. (1990) Morphology of the differentiation and maturation of LLC-PK1 epithelia. *J. Cell Physiol.*, **142** (2), 247–254.

69 Handler, J.S., Perkins, F.M., and Johnson, J.P. (1980) Studies of renal cell function using cell culture techniques. *Am. J. Physiol.*, **238** (1), F1–F9.

70 Clarke, H., Soler, A.P., and Mullin, J.M. (2000) Protein kinase C activation leads to dephosphorylation of occludin and tight junction permeability increase in LLC-PK1 epithelial cell sheets. *J. Cell Sci.*, **113** (Pt 18), 3187–3196.

71 Prozialeck, W.C. *et al.* (2006) Epithelial barrier characteristics and expression of cell adhesion molecules in proximal tubule-derived cell lines commonly used for *in vitro* toxicity studies. *Toxicol. In Vitro*, **20** (6), 942–53.

72 Alexandre, M.D., Lu, Q., and Chen, Y.H. (2005) Overexpression of claudin-7 decreases the paracellular Cl^- conductance and increases the paracellular Na^+ conductance in LLC-PK1 cells. *J. Cell Sci.*, **118** (Pt 12), 2683–93.

73 Hori, R. *et al.* (1993) Transport of organic anion in the OK kidney epithelial cell line. *Am. J. Physiol.*, **264** (6 Pt 2), F975–F980.

74 Bringhurst, F.R. *et al.* (1993) Cloned, stably expressed parathyroid hormone (PTH)/PTH-related peptide receptors activate multiple messenger signals and biological responses in LLC-PK1 kidney cells. *Endocrinology*, **132** (5), 2090–2098.

75 de Larco, J.E. and Todaro, G.J. (1978) Epithelioid and fibroblastic rat kidney cell clones: epidermal growth factor (EGF) receptors and the effect of mouse sarcoma virus transformation. *J. Cell Physiol.*, **94** (3), 335–342.

76 Limonciel, A. *et al.* (2012) Oxidative stress induced by potassium bromate exposure results in altered tight junction protein expression in renal proximal tubule cells. *Arch. Toxicol.*, **86** (11), 1741–1751.

77 Racusen, L.C. *et al.* (1997) Cell lines with extended *in vitro* growth potential from human renal proximal tubule: characterization, response to inducers, and comparison with established cell lines. *J. Lab. Clin. Med.*, **129** (3), 318–29.

78 Ratajczak, M.Z. *et al.* (2012) Pluripotent and multipotent stem cells in adult tissues. *Adv. Med. Sci.*, **57** (1), 1–17.

79 Hendry, C.E. and Little, M.H. (2012) Reprogramming the kidney: a novel

approach for regeneration. *Kidney Int.*, **82** (2), 138–146.
80 Bussolati, B. *et al.* (2005) Isolation of renal progenitor cells from adult human kidney. *Am. J. Pathol.*, **166** (2), 545–55.
81 Oliver, J.A. *et al.* (2004) The renal papilla is a niche for adult kidney stem cells. *J. Clin. Invest.*, **114** (6), 795–804.
82 Romagnani, P. (2009) Toward the identification of a "renopoietic system"? *Stem Cells*, **27** (9), 2247–2253.
83 Bonventre, J.V. (2003) Dedifferentiation and proliferation of surviving epithelial cells in acute renal failure. *J. Am. Soc. Nephrol.*, **14** (Suppl. 1), S55–S61.
84 Humphreys, B.D. *et al.* (2011) Repair of injured proximal tubule does not involve specialized progenitors. *Proc. Natl. Acad. Sci. USA*, **108** (22), 9226–9231.
85 Kale, S. *et al.* (2003) Bone marrow stem cells contribute to repair of the ischemically injured renal tubule. *J. Clin. Invest.*, **112** (1), 42–49.
86 Lange, C. *et al.* (2005) Administered mesenchymal stem cells enhance recovery from ischemia/reperfusion-induced acute renal failure in rats. *Kidney Int.*, **68** (4), 1613–1617.
87 Morigi, M. *et al.* (2004) Mesenchymal stem cells are renotropic, helping to repair the kidney and improve function in acute renal failure. *J. Am. Soc. Nephrol.*, **15** (7), 1794–1804.
88 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126** (4), 663–676.
89 Takahashi, K. *et al.* (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131** (5), 861–872.
90 Zhou, T. *et al.* (2011) Generation of induced pluripotent stem cells from urine. *J. Am. Soc. Nephrol.*, **22** (7), 1221–1228.
91 Gadue, P. *et al.* (2006) Wnt and TGF-beta signaling are required for the induction of an *in vitro* model of primitive streak formation using embryonic stem cells. *Proc. Natl. Acad. Sci. USA*, **103** (45), 16806–16811.
92 Soares, M.L. *et al.* (2005) Functional studies of signaling pathways in peri-implantation development of the mouse embryo by RNAi. *BMC Dev. Biol.*, **5**, 28.
93 Carlson, B.M. (2013) *Human Embryology and Developmental Biology*, 5th edn, WB Saunders Company.
94 James, R.G. *et al.* (2006) Odd-skipped related 1 is required for development of the metanephric kidney and regulates formation and differentiation of kidney precursor cells. *Development*, **133** (15), 2995–3004.
95 Cartry, J. *et al.* (2006) Retinoic acid signalling is required for specification of pronephric cell fate. *Dev. Biol.*, **299** (1), 35–51.
96 Dressler, G.R. (2006) The cellular basis of kidney development. *Annu. Rev. Cell Dev. Biol.*, **22**, 509–529.
97 Davies, J. (2001) Intracellular and extracellular regulation of ureteric bud morphogenesis. *J. Anat.*, **198** (Pt 3), 257–264.
98 Dudley, A.T., Godin, R.E., and Robertson, E.J. (1999) Interaction between FGF and BMP signaling pathways regulates development of metanephric mesenchyme. *Genes Dev.*, **13** (12), 1601–1613.
99 Tsang, T.E. *et al.* (2000) Lim1 activity is required for intermediate mesoderm differentiation in the mouse embryo. *Dev. Biol.*, **223** (1), 77–90.
100 Costantini, F. and Shakya, R. (2006) GDNF/Ret signaling and the development of the kidney. *Bioessays*, **28** (2), 117–127.
101 Wellik, D.M., Hawkes, P.J., and Capecchi, M.R. (2002) Hox11 paralogous genes are essential for metanephric kidney induction. *Genes Dev.*, **16** (11), 1423–1432.
102 Xu, P.X. *et al.* (2003) Six1 is required for the early organogenesis of mammalian kidney. *Development*, **130** (14), 3085–3094.
103 Bruce, S.J. *et al.* (2007) *In vitro* differentiation of murine embryonic stem cells toward a renal lineage. *Differentiation*, **75** (5), 337–349.
104 Kim, D. and Dressler, G.R. (2005) Nephrogenic factors promote

differentiation of mouse embryonic stem cells into renal epithelia. *J. Am. Soc. Nephrol.*, **16** (12), 3527–3534.

105 Vigneau, C. *et al.* (2007) Mouse embryonic stem cell-derived embryoid bodies generate progenitors that integrate long term into renal proximal tubules *in vivo*. *J. Am. Soc. Nephrol.*, **18** (6), 1709–1720.

106 Takasato, M. *et al.* (2014) Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney. *Nat. Cell Biol.*, **16** (1), 118–126.

107 Xia, Y. *et al.* (2013) Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells. *Nat. Cell Biol.*, **15** (12), 1507–1515.

108 Narayanan, K. *et al.* (2013) Human embryonic stem cells differentiate into functional renal proximal tubular-like cells. *Kidney Int.*, **83** (4), 593–603.

109 Song, B. *et al.* (2012) The directed differentiation of human iPS cells into kidney podocytes. *PLoS One*, **7** (9), e46453.

110 Briere, N., Ferrari, J., and Chailler, P. (1991) Insulin and transferrin restore important cellular functions of human fetal kidney in serum-free organ culture. *Biochem. Cell Biol.*, **69** (4), 256–262.

111 Bertrand, L. and Briere, N. (1989) Effect of hydrocortisone on the maturation of human foetal kidney explants in serum-free organ culture. *Biochem. Cell Biol.*, **67** (2–3), 121–127.

112 Avner, E.D. *et al.* (1985) Growth factor requirements of organogenesis in serum-free metanephric organ culture. *In Vitro Cell Dev. Biol.*, **21** (5), 297–304.

113 Van der Biest, I. *et al.* (1994) Characterization of pure proximal and heterogeneous distal human tubular cells in culture. *Kidney Int.*, **45** (1), 85–94.

114 Sikka, P.K. and McMartin, K.E. (1996) Normal rat kidney proximal tubule cells in primary and multiple subcultures. *In Vitro Cell Dev. Biol. Anim.*, **32** (5), 285–291.

115 Jennings, P. *et al.* (2009) Inter-laboratory comparison of human renal proximal tubule (HK-2) transcriptome alterations due to cyclosporine A exposure and medium exhaustion. *Toxicol. In Vitro*, **23** (3), 486–499.

116 Ludwig, T.E. *et al.* (2006) Feeder-independent culture of human embryonic stem cells. *Nat. Methods*, **3** (8), 637–646.

117 Pettersen, E.O. *et al.* (2005) Pericellular oxygen depletion during ordinary tissue culturing, measured with oxygen microsensors. *Cell Prolif.*, **38** (4), 257–267.

118 Gstraunthaler, G., Seppi, T., and Pfaller, W. (1999) Impact of culture conditions, culture media volumes, and glucose content on metabolic properties of renal epithelial cell cultures. Are renal cells in tissue culture hypoxic? *Cell Physiol. Biochem.*, **9** (3), 150–172.

119 Bernhardt, W.M. *et al.* (2006) Expression of hypoxia-inducible transcription factors in developing human and rat kidneys. *Kidney Int.*, **69** (1), 114–122.

120 Felder, E. *et al.* (2002) LLC-PK(1) cells maintained in a new perfusion cell culture system exhibit an improved oxidative metabolism. *Cell Physiol. Biochem.*, **12** (2–3), 153–162.

121 Koppelstaetter, C. *et al.* (2004) Assessment of a new cell culture perfusion apparatus for *in vitro* chronic toxicity testing. Part 1: technical description. *ALTEX*, **21** (2), 51–60.

122 Schimek, K. *et al.* (2013) Integrating biological vasculature into a multi-organ-chip microsystem. *Lab Chip*, **13** (18), 3588–3598.

123 Jang, K.J. *et al.* (2013) Human kidney proximal tubule-on-a-chip for drug transport and nephrotoxicity assessment. *Integr. Biol.*, **5** (9), 1119–1129.

124 Gunness, P. *et al.* (2013) 3D organotypic cultures of human HepaRG cells: a tool for *in vitro* toxicity studies. *Toxicol. Sci.*, **133** (1), 67–78.

125 Honegger, P. *et al.* (2011) Preparation, maintenance, and use of serum-free aggregating brain cell cultures. *Methods Mol. Biol.*, **758**, 81–97.

126 Zurich, M.G. *et al.* (2013) Evaluation of aggregating brain cell cultures for the detection of acute organ-specific toxicity. *Toxicol. In Vitro*, **27** (4), 1416–1424.

127 DesRochers, T.M. *et al.* (2013) Bioengineered 3D human kidney tissue, a platform for the determination of nephrotoxicity. *PLoS One*, **8** (3), e59219.

128 Handler, J.S., Preston, A.S., and Steele, R. E. (1984) Factors affecting the differentiation of epithelial transport and responsiveness to hormones. *Fed. Proc.*, **43** (8), 2221–2224.

129 Zlatkine, P. *et al.* (1993) Transport of choline by Madin–Darby canine kidney cells. *Biochim. Biophys. Acta*, **1153** (2), 237–242.

130 Bijuklic, K. *et al.* (2007) Migration of leukocytes across an endothelium–epithelium bilayer as a model of renal interstitial inflammation. *Am. J. Physiol. Cell Physiol.*, **293** (1), C486–C492.

131 Pfaller, W. and Gstraunthaler, G. (1998) Nephrotoxicity testing *in vitro* – what we know and what we need to know. *Environ. Health Perspect.*, **106** (Suppl. 2), 559–569.

132 Limonciel, A. and Jennings, P. (2014) A review of the evidence that ochratoxin A is an Nrf2 inhibitor: implications for nephrotoxicity and renal carcinogenicity. *Toxins*, **6** (1), 371–379.

133 Jennings, P. *et al.* (2007) Cyclosporine A induces senescence in renal tubular epithelial cells. *Am. J. Physiol. Renal Physiol.*, **293** (3), F831–F838.

134 Jennings, P. *et al.* (2013) An overview of transcriptional regulation in response to toxicological insult. *Arch. Toxicol.*, **87** (1), 49–72.

135 Crean, D. *et al.* (2012) Glucose reintroduction triggers the activation of Nrf2 during experimental ischemia reperfusion. *Mol. Cell. Biochem.*, **366** (1–2), 231–238.

136 Wilmes, A. *et al.* (2011) Identification and dissection of the Nrf2 mediated oxidative stress pathway in human renal proximal tubule toxicity. *Toxicol. In Vitro*, **25** (3), 613–622.

11
The Zebrafish Model in Toxicology

Natalie Mesens

11.1
The Need for a Physiologically Relevant Organ Model in Drug Toxicity Testing

The current challenges facing the pharmaceutical industry are unprecedented in its history. Whereas it has had a significant positive impact on the health and longevity of humans at the end of the twentieth century, nowadays, in the early twenty-first century, the pharmaceutical industry is pressurized by increasing costs and timelines of drug development, while the number of drugs that reach the market is decreasing. Approval rates of NCEs (new chemical entities, new drugs) by the US FDA are lower than those at any other time. Recent data indicate that the average success rate for all therapeutic areas is approximately 11%, or only one in nine compounds makes it through development [1]. This large attrition in development not only jeopardizes the ability to address unmet medical need, but also often compromises human safety. A recent investigation examines the root causes of compound attrition in the clinical development [1]. In 1991, adverse pharmacokinetic and bioavailability results were the most significant cause of attrition, and accounted for 40% of all attrition. By 2000, these factors had dramatically reduced as a cause of attrition in drug development, and contributed less than 10%. The major causes of attrition in the clinic in 2000 were lack of efficacy (accounting for approximately 30% of the failures) and safety (toxicology and clinical safety accounting for a further approximately 30%). On the one hand, preclinical attrition due to toxicity is high [2]. On the other hand, a considerable number of drug candidates fail because of safety problems in humans in the clinical phase of evaluation despite having passed the entire toxicity testing program [1]. A recent survey identified three major toxicities as the main clinical adverse events reported: hepatic toxicity (14%), cardiovascular toxicity (16%), and neurological toxicity (22%) [3]. The increasing rate of failures due to toxicity demonstrates an urgent need for more predictive models. Identifying and eliminating drug candidates with toxic liabilities earlier from the development cycle is a preferred strategy. The pharmaceutical industry tries to take advantage of the molecular knowledge generated over the past few years underlying organ toxicities by developing high-throughput

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

in vitro screening assays for detecting these pathways of toxicity. Whereas great improvements have been made in the past few years, the predictive capacity of these *in vitro* assays toward true organ toxicity remains poor and the extrapolation from cytotoxicity *in vitro* toward risk of developing human drug-induced organ toxicity continues to be a challenging task. In a way of avoiding unnecessary use and cost of performing mammalian studies early in development, a possible solution to the limitations of cell-based screening is the use of a small animal species in the screening strategy. As genes, receptors, and molecular processes are highly conserved across animal phyla, studies with other animal species could be representative for higher, more complex animals [4]. Examples of model invertebrate organisms are *Drosophila melanogaster* and *Caenorhabditis elegans*, while model vertebrate organisms include *Danio rerio* (zebrafish). Of these small, well-characterized surrogate animals, *D. rerio* is preferred because its cardiovascular, nervous, and digestive systems are similar to the mammalian counterparts at the anatomical, physiological, and molecular levels [5].

Zebrafish have several inherent advantages for drug screening: due to the small size of the embryo, the amount of compound required for toxicity assessment is much less compared with mammalian models. This is important in view of the fact that identified lead compounds are usually available in limited amount. In addition to its small size, the availability of a large number of zebrafish embryos allows toxicity screening to be performed in microwell plates; this makes the system amenable to automation in a high-throughput manner [6,7]. It is a major advantage since the low throughput of mammalian models creates a major bottleneck to assessing the numerous "hits" identified from *in vitro* screening [8]. Zebrafish larvae might thus fill the gap between high-throughput *in vitro* screening and low-throughput mammalian toxicity testing (Figure 11.1). Zebrafish embryos allow toxicity to be assessed in the context of a complex dynamic growing organism. This may be an important consideration when compared with cell-based models such as embryonic stem cell models advocated for *in vitro* toxicity screening [9]. This is especially important in the context of pharmacology and toxicology when many of the potentially targeted organ systems, for example, the nervous, cardiac, digestive (liver in particular), immune, musculoskeletal, vasculature, and kidney, are functional in the late embryonic and early larval stages of the zebrafish. This has further led to the proposal of the use of zebrafish as a model for assessing off-target drug effects, which in recent times is recognized as a cause of drug attrition [10]. According to the current European Directive 2010/63/EU [11], the zebrafish larva is viewed as an embryo or eleutheroembryo and is considered as an *in vitro* model until it reaches the state of exogenous feeding. Moreover, the availability of many transgenic lines with fluorescing organs/tissues developed by introduction of fluorescent (e.g., green fluorescent protein or GFP) reporter genes under a tissue-specific promoter allows real-time monitoring and high-resolution qualitative and quantitative assessments of specific organ/tissue toxicity [12].

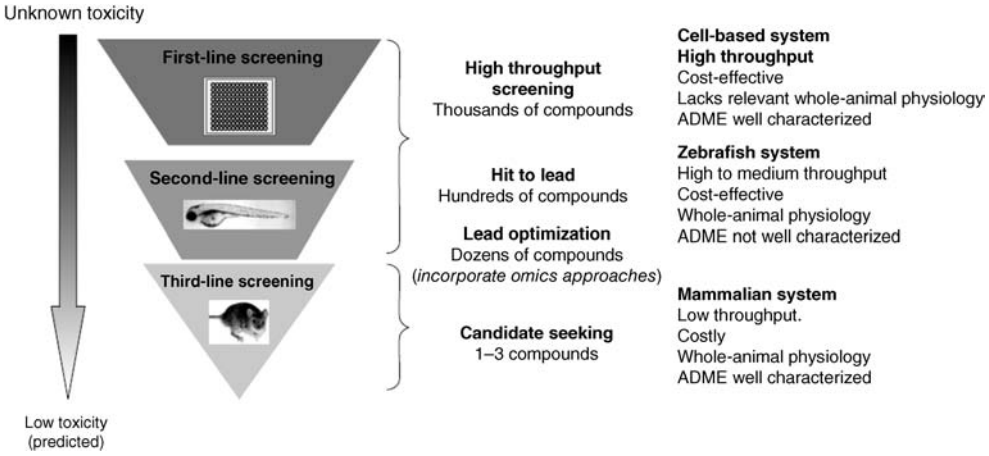


Figure 11.1 Incorporating zebrafish into preclinical toxicity screening. Zebrafish could enhance preclinical toxicity screening by its strategic placement between *in vitro* cell-based and mammalian models along the drug development pipeline. It can be employed for toxicity screening in early stages along the development pipeline. This could alert researchers of potential hazards and help remove candidate compounds with significant safety liabilities at earlier stages, thereby reducing drug attrition costs. (From Ref. [8].)

11.2
Extensive Knowledge about Genetics, Development, and Physiology of *D. rerio*

The zebrafish has long been used as an experimental model to study chemical toxicity ranging from mutagens, carcinogens, teratogens to direct toxicants since 1950s [13]. One reason for the long history is that fish acute toxicity testing has been an important, if not mandatory, component in the base set of data requirements for ecotoxicity testing [14], which involved several fish species including zebrafish. From 1980s to mid-1990s, the zebrafish became a premier vertebrate model for developmental and genetic studies, and within the next decade it positioned itself as a biomedical model for various human disorders that could aid in discovering novel therapeutics. This led to a number of studies employing zebrafish for drug discovery. The large genomic resources, availability of "omics" platforms, and amenability to various molecular and bioimaging techniques provide versatility of investigating toxicity at different levels. Toxicity in zebrafish can be investigated from the morphological to the molecular levels, from a single biomolecule to hundreds of them simultaneously, and from a single cell type or organ to multiorgan systems including the whole organism. The availability of zebrafish models of human diseases that could aid in drug discovery makes it more attractive for drug toxicity screening. This is because pathology, pharmacology, and toxicology can now be performed within the same organism, facilitating tractability, comparison, and connectivity of data and findings between these experiments.

11.3
Studies of Specific Organ Toxicities in Zebrafish Embryos and Larvae

11.3.1
Cardiotoxicity

Cardiac repolarization abnormalities, often leading to life-threatening arrhythmias, represent one of the most important types of drug toxicity [15–17]. Drugs that cause clinical arrhythmias are often associated with prolongation of the QT interval measured on the electrocardiogram. The majority of drugs that result in clinical problems interfere with repolarization by blocking the rapid component of the delayed rectifier potassium current (I_{Kr}). This can be captured in an *in vitro* assay (hERG assay); however, *in vitro* hERG inhibition has been limited in its predictive value [18–21]. The lack of utility of current assays and the failure of these assays to detect drug effects on other targets that may modulate hERG indirectly are a result of the intrinsic complexity of the myocardial substrate and the involvement of extracardiac factors in the genesis of actual toxic arrhythmic events [22–25]. Partly because of the feasibility of characterizing the toxicity at every level – from the individual ion channels through to integrated physiology – this area has been one of the most extensively studied in zebrafish toxicology.

In zebrafish, the heart is the first organ to develop and function and a beating heart forms by 22 h postfertilization (hpf). By 48 hpf, the cardiovascular system is fully functional and exhibits a complex repertoire of ion channels and metabolic processes [26]. Zebrafish ERG is expressed in the early stages of zebrafish development and the amino acid sequences of the pore-forming domain of the zebrafish ERG and human ERG are 99% conserved [27]. Although zebrafish and mammalian hearts differ in structure and zebrafish lack a pulmonary system, they exhibit similar functional characteristics, including (1) blood flow from a major vein into an atrium, (2) blood moves through a muscular ventricle for delivery to the aorta, (3) valves direct blood flow, (4) a specialized endocardium musculature drives a high-pressure system, (5) an electrical system regulates rhythm, and (6) heartbeat is associated with pacemaker activity [27]. Knocking down the zERG gene in the zebrafish (hERG ortholog) results in a characteristic arrhythmia, with two atrial beats coupled to each ventricular beat. Known QT-prolonging drugs when tested in 3 days postfertilization (dpf) embryos cause this specific arrhythmia in a concentration-dependent manner with lower concentrations inducing bradycardia and higher concentrations leading to 2 : 1 decoupling followed by, in some cases, a more pronounced decoupling, irregular arrhythmia, or a complete ventricular block [27,28]. This atrial–ventricular decoupling effect was able to detect QT-prolonging drugs known to block hERG as well as indirect QT-prolonging drugs, suggesting that this may be taken as a surrogate signaling QT prolongation [29]. Using these simple heart rate responses, investigators have been able to establish an excellent correlation with known adult human repolarization cardiac toxicity and recapitulate clinically relevant drug–drug

interactions [29]. Even dramatic effects on cardiac function are well tolerated by the larvae, which can survive for 4–5 days without an active circulation.

Measurements of heartbeat (number of atrial and ventricular contractions) over a period of 15–20 s are determined either manually (by eye) [30] or by using customized software applications. In the latter scenario, video imaging using transmitted light with wild-type larvae [27] or fluorescence imaging with transgenic larvae expressing green fluorescent protein in the myocardium [31] has been reported. Both software applications offer relatively high-throughput readouts by identifying changes in average pixel density in the region of the atrium and ventricle, and when plotted over time, the number of contractions for each chamber and average heart rate are automatically determined. It also proved feasible to define a range of sophisticated second-tier assays with higher resolution and lower throughput that have been used in series to optimize the overall sensitivity and specificity of the approach. These assays include calcium imaging and optical voltage mapping that enable characterization of integrated myocardial electrophysiology at a resolution comparable to that in current "state-of-the-art" canine or rabbit models. Using the zebrafish, investigators have also been able to define action potential prolongation in the setting of QT-prolonging drugs, known ion channel mutations, or novel genes recently implicated in cardiac repolarization.

Currently published studies show promising results in predicting dangerous QT prolongation [11,28]. Using the same assays, a blinded screen of FDA-approved compounds predicted 20 of 23 known agents that cause the drug-induced arrhythmia (*torsades de pointes*) and also successfully reproduced a broad range of relevant drug–drug interactions; however, a number of false negatives were reported, mainly due to poor absorption [29]. The recapitulation of these canonical multiorgan interactions, which are inaccessible *in vitro*, demonstrates a major advantage of *in vivo* modeling [32]. As noted, these data suggest robust functional conservation between humans and zebrafish at multiple levels, and they have formed the basis of commercial assays for cardiotoxic drugs using screening technologies in larval fish. Similarly, panels of drugs known to perturb contractility and/or vasomotor tone have been tested and found to recapitulate very accurately the physiological effects observed with the same agents in humans [33].

11.3.2 Neurotoxicity

Drug-induced neurological effects are remarkably common and are among the most frequent reasons for poor drug compliance, even in the context of well-established benefits. The sheer complexity of the nervous system makes *in vivo* modeling attractive, and several groups have pioneered neurological assays in the zebrafish. The overall organization of the zebrafish brain is similar to other vertebrates, having matched defined areas such as the hypothalamus and olfactory bulb, encompassing structures of the lateral pallium, which appear to be homologous to the mammalian hippocampus [34]. In addition, the main neurotransmitter systems such as the cholinergic, 5-hydroxytryptaminergic,

dopaminergic, and noradrenergic pathways are also present and have been mapped throughout the brain [35,36]. Zebrafish larvae have recently become the focus of neurobehavioral studies as they display learning, sleep, drug addiction, and neurobehavioral phenotypes that are related to those seen in humans [37]. Transgenic lines are created to study neurological disorders such as Alzheimer's disease [38–40] and Huntington's disease [41]. Similarly, genetic and chemical effects on each of these behaviors in humans can be recapitulated in the zebrafish. However, large-scale chemical screens exploiting these assays have only just begun to be undertaken [42]. An alternative approach, focusing on a single neurological circuit, but adapting this for much higher throughput, has led to the characterization of the neurological effects of tens of hundreds of small molecules [43]. This strategy suggests not only that effectively clustering neuroactive drugs using robust quantitative metrics is possible, but also that such clustering will extend to potential neurotoxicity. Ongoing work in screens of truly integrated behaviors will shed additional light on the utility of the zebrafish for the study of neurotoxicity.

From a safety pharmacology perspective, drug candidates are screened for CNS side effects, which include seizure liability and sedative or stimulant effects. A number of sedatives were tested in the zebrafish and caused hypomotility, in concordance with the effects seen in humans [37]. Recently, Winter *et al.* [44] conducted a convulsant activity assay in which 25 compounds were tested on their ability to increase the swimming pattern of zebrafish. Thirteen of the seventeen convulsants were positive in the zebrafish locomotor assay and five of eight negative compounds were negative. The authors suggested that many of the false-negative compounds were not absorbed well by the zebrafish, but this possibility was not investigated in depth. The false positives are of greater concern. This might be due to the fact that the blood–brain barrier of the larval zebrafish is not fully formed until 10 dpf. Therefore, the assessment of CNS-mediated effects in larvae may erroneously identify compounds that are excluded from the brain in older fish and in mammals [45].

11.3.3 Hepatotoxicity

Drug-induced liver injury (DILI) accounts for approximately 11–13% of acute liver failure cases in the United States and is the most common cause of death related to this condition [46,47]. Although DILI is implicated in acute liver failure and results in black-box warnings, withdrawal from the marketplace, or restriction of its use after marketing approval has been obtained [48,49], many agents cause DILI during preclinical or clinical development that never make it to the marketplace. Conventional cytotoxicity assays have been used for the detection of hepatotoxicants [50], but can be very limited in their sensitivity to detect hepatotoxic liability because they typically utilize a single cell type. Although screening for specific hepatic pathologies, such as steatosis, can also be employed [51], compounds that cause injury through more complex

mechanisms (e.g., immune-based injury or those of an idiosyncratic nature) are often missed as a result of the simplicity of the cell model systems being used early to screen compounds. Despite these known limitations, recent advances in cell-based screening models have shown good sensitivity and specificity using high-content screening (HCS) methods that employ endpoints relevant to mechanisms of hepatocellular injury [52]. The sensitivity of these cell-based models can be further increased when cultured hepatocytes are incubated with inflammatory cytokines concurrently with the test agent [53,54]. These results illustrate the power of simulating the complex microenvironment of the liver even *in vitro*, rather than the "standardized" tissue culture environment. *In vivo* models may also permit probing the microenvironment of the "primed" liver. Several specific animal models (e.g., endotoxin-stimulated hepatotoxicity model) [55] and the heterozygous $Sod2^{+/-}$ mouse [56,57] have been purported to identify hepatotoxicants of clinical concern. However, these models have mostly been employed in retrospective investigations, rather than prospectively, as an *in vivo* screening model early in drug discovery. These are mostly the result of impractical drug supply requirements and associated higher costs of animal testing. As such, relatively less expensive alternative whole animal models, such as the zebrafish, are currently being investigated.

The zebrafish liver develops from anterior endodermal progenitor cells and is identifiable by 22 hpf. Expression of genes indicative of hepatocyte function is first detectable by 32 hpf, and hepatic outgrowth begins at 72 phf. At 5 dpf, bile production, serum protein secretion, glycogen storage, lipogenesis, and xenobiotic metabolism are fully operational [58]. The teleost liver has distinguishable histological characteristics compared with that in mammals. The portal veins, hepatic arteries, and large biliary ducts are distributed stochastically within the hepatic parenchyma but are not grouped in portal tracts as in the mammals. Hepatocytes are arranged as tubules that enclose small bile ducts rather than as bilayered hepatocyte plates in the human liver. The intrahepatic bile ducts are derived from the bile canaliculi and form a network that is collected in the gall bladder through large ducts and an extrahepatic biliary system [59,60]. Despite cellular and anatomical architectural differences, hepatocytes and biliary ductules maintain the same functions. Mutations have identified mutants with features of liver diseases and a model for hepatic steatosis, choledochal cysts, and cholestasis has been established [61]. Treatment with carcinogens generated liver tumors [62] and γ-hexachlorocyclohexane, thioacetamide, and alcohol can induce hepatic steatosis [63–65]. CYP3A4 and CYP2D6 enzymes, catalyzing the majority of known drug metabolizing reactions, have zebrafish orthologs. CYP3A4 and CYP2D6 functional activity assays have been performed in zebrafish using human CYP-specific substrates and similar responses have been identified, indicating a potential similar metabolism in zebrafish and humans [66]. Because zebrafish larvae are virtually transparent, the liver of a 120–144 hpf larva can be viewed dorsolaterally anterior to the gut and posterior to the pericardium with simply a stereomicroscope and transmitted light. At this age, the liver is perfused with circulating blood cells, is fairly globular in structure, and has a

clearly recognizable periphery against the neighboring tissues [67]. Initial hepatotoxicity assays were created to assess morphological endpoints of toxicity after treatment with different compounds, which, unlike mammalian studies, had no reliance on dissection and histology. However, it should be noted that when viewed in transverse section, hepatic tissue can appear to surround the esophagus, so each larva needs to be assessed in various orientations to appropriately screen the whole liver [68]. At 120 hpf (after a 48 h incubation), larvae were anesthetized in MS-222 (tricaine) and screened for three specific phenotypic endpoints of hepatotoxicity: changes in liver size (i.e., hepatomegaly), liver abnormality (i.e., tissue degradation), and yolk retention (an endpoint for liver dysfunction, as yolk is utilized through the liver and utilization is diminished if the liver is impaired) (Figure 11.2). As an example, the typical phenotype induced by a hepatotoxicant (i.e., changes in the texture and color of the liver were documented with tissue becoming amorphous and darker than controls with a brown and gray hue) has been reported previously for brefeldin A [69] and certain zebrafish liver mutants [59]. Two independent studies revealed good predictive capacity of this assay [68,70] toward known hepatotoxicants; however, the advantages to a single cell-based assay could not be identified [68]. In addition, He *et al.* [70] showed that the drug-induced liver degeneration, liver size change, and yolk sac retention could be quantitatively assessed using an image-based morphometric analysis. Furthermore, to confirm that the visual observation of hepatotoxicity induced truly represents pathological liver damage, larvae treated with acetaminophen (APAP), aspirin, tetracycline HCl, sodium valproate, cyclophosphamide, and erythromycin were assessed histologically and all six hepatotoxic drugs induced zebrafish liver degeneration. As demonstrated in Figure 11.3, liver from zebrafish treated with APAP showed loose cell-to-cell contact and the cells were dissociated and irregular in shape with various grades of tiny and large vacuoles, and the number of hepatic parenchyma cells was decreased. True hepatocellular necrosis, however, was absent. Further histological confirmation of major mechanisms of drug-induced liver injury was performed by Driessen *et al.* [71] through histological analysis after exposure to prototypical hepatotoxicants inducing cholestasis, steatosis, and apoptosis. Histological evidence of cholestasis was absent, steatosis was identified as the appearance of large vacuoles, but also appeared during treatment of a cholestasis inducer, and true apoptosis was also absent. On the other hand, marks of cell death, particularly chromatin condensation and cytoplasmic eosinophilia, were observed, although more compound than class specific. Additional observations were chromophobic and eosinophilic vacuolization. In summary, the final histopathology of hepatotoxic effects was likely related to life stage-dependent capability of hepatocytes, and hepatotoxic responses were less specific as anticipated. Microarray analysis of whole larvae also revealed a diversity of changed pathways, but liver specificity was difficult to prove due to the potential interference of other organ effects.

The above studies shed light not only on the possibilities but also on the complexity of the zebrafish larvae as a screening model for assessing drug-induced

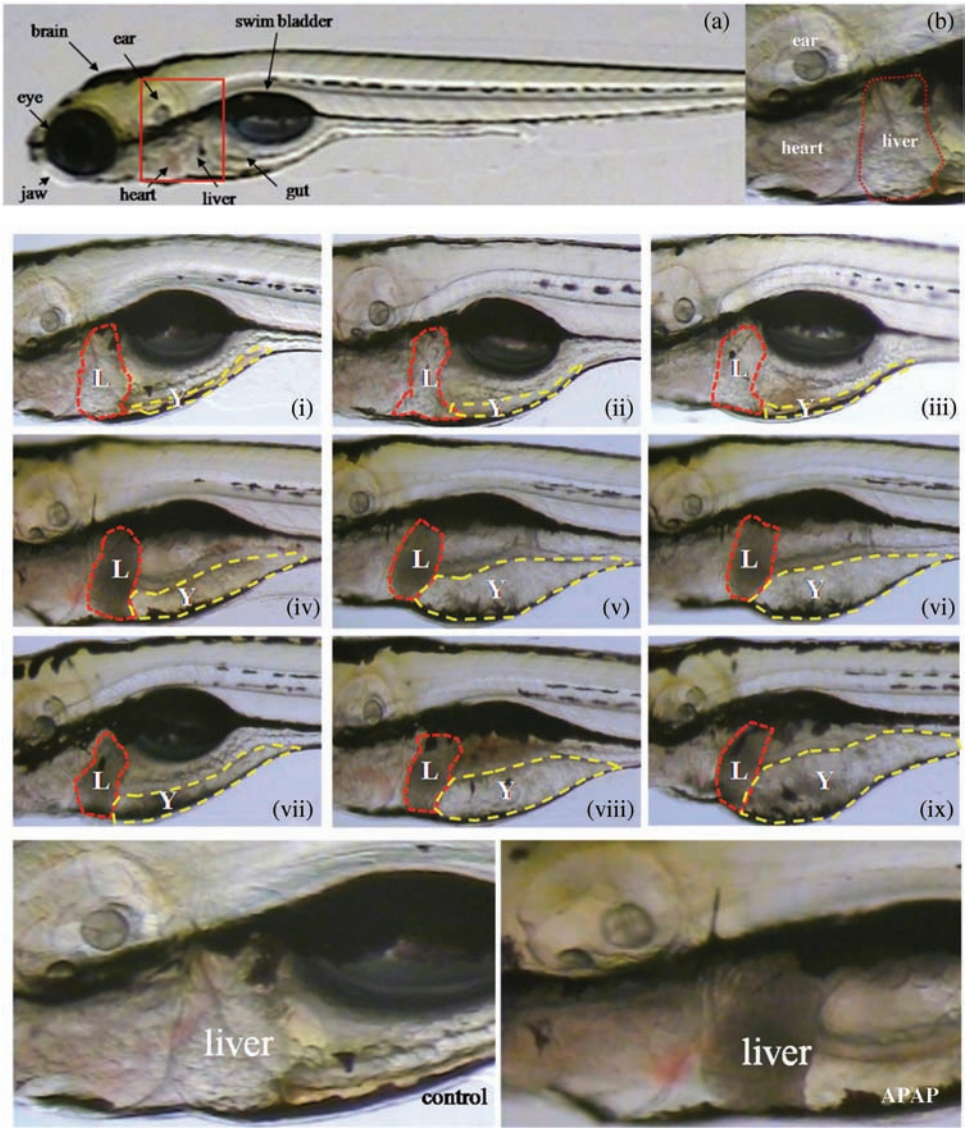


Figure 11.2 Visual phenotype of hepatotoxicity in zebrafish at 120 hpf after exposure to drugs/compounds from 72 to 120 hpf (a). Vehicle-treated larval zebrafish exhibited a clear, healthy liver (i). Larval zebrafish treated with mammalian hepatotoxic drugs exhibited liver tissue degradation, hepatatrophia, and yolk sac retention (iv–ix). Two mammalian nonhepatotoxic compounds sucrose and biotin did not show any adverse effect on zebrafish liver (ii, iii). Part (b) shows the liver sections of representative figures indicating normal liver in control zebrafish and liver degeneration in zebrafish treated with APAP. L: liver; Y: yolk sac; (i) vehicle control (0.1% DMSO); (ii) biotin; (iii) sucrose; (iv) APAP; (v) aspirin; (vi) tetracycline HCl; (vii) valproate sodium; (viii) cyclophosphamide; (ix) erythromycin. (From Ref. [70].)

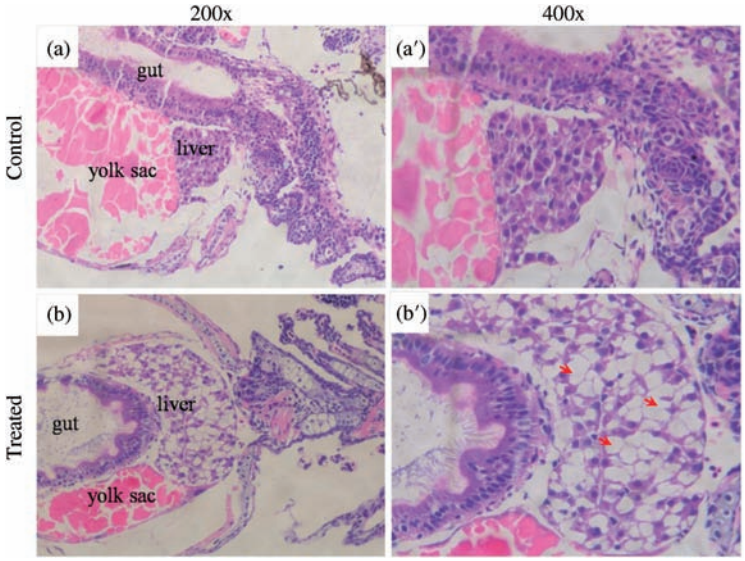


Figure 11.3 Representative histological pictures of larval zebrafish at 120 hpf after treatment with testing drugs. Pictures labeled as (a) and (a′) were the liver histopathology from vehicle (0.1% DMSO)-treated zebrafish and the pictures labeled as (b) and (b′) were the liver histopathology from zebrafish treated with a mammalian hepatotoxic drug APAP. (From Ref. [70].)

hepatotoxicity. In our hands, we obtained the same preliminary conclusions. We developed a hepatotoxicity screening assay in transgenic zebrafish larvae expressing a liver-specific fatty acid binding protein (fabp10)-driven dsRED protein. Analysis of hepatotoxic effects can then easily be assessed using high-throughput imaging by analyzing the reduction or increase in fluorescence by automated image analysis of the treated larvae, allowing semi-high-throughput assessment of DILI in drug discovery. However, due to the current lack of concordance between the responses of the embryonic immature zebrafish liver and a mammalian liver, extensive validation of hepatotoxicity screening assays with prototypical DILI inducers will be necessary in the future to gain confidence in the power of zebrafish larvae for predicting human DILI.

11.3.4 Teratogenicity

According to regulatory guidelines, each drug in development for administration to women of child-bearing potential must be tested for embryotoxic and teratogenic potential in a rodent and a nonrodent species. Depending on the phase within the drug development process, the indication, and the related target group, the drug development plan may require limited or more extensive animal data in order to move forward [72,73]. Ultimately, achieving registration of a

drug translates into the use of a large number of animals, generally rats and rabbits, for the teratogenicity safety assessment [74], which is not in line with the current trend of applying the 3R principle, that is, replacement, reduction, or refinement of animal use, in general. Looking at alternative testing methods, the larval zebrafish seems to be a promising model for the evaluation of adverse drug reactions and a zebrafish developmental toxicity assay could be used to identify compounds harmful to embryofetal development early in drug development.

Zebrafish development has been well characterized [75]. The eggs remain transparent from fertilization up to and beyond pharyngulation when the tissues become dense and pigmentation is initiated. This allows unobstructed observation of the main morphological changes during earlier developmental stages. Furthermore, zebrafish embryos that are malformed, lack organs, or display organ dysfunction can usually survive well beyond the time at which those organs normally start to function in healthy individuals [4]. In addition, fish are sensitive to chemical exposure during early development [76]. These characteristics make the zebrafish an attractive candidate for screening of teratogens and the elucidation of mechanisms thereof [4]. As an example, the primary target of the strong human teratogen thalidomide was recently elucidated by the use of zebrafish larvae [77].

The teratogenic potential of compounds can then be predicted quantitatively by ranking zebrafish embryos based on a scoring system for phenotypic changes that is conceptually similar to morphological assessments conducted using *in vivo* embryofetal development of mammals (Table 11.1; Figures 11.4 and 11.5) [78,79].

A number of groups have reported exploratory studies with the zebrafish [9,82–85]. Although different incubation periods were used and different endpoints for teratogenicity were scored, overall zebrafish teratogenicity assays have been shown to correctly classify *in vivo* teratogens and nonteratogens with overall concordance of 72–92% [86]. A recently developed zebrafish teratogenicity assay was shown to have successfully categorized 87% of the 31 test compounds as *in vivo* teratogens or nonteratogens, with only 2 false positives and 2 false negatives [80]. A meeting convened to address the use of zebrafish and other *in vitro* models (i.e., embryonic stem cells and rodent whole-embryo culture) of developmental toxicity found that zebrafish has the potential to provide a level of predictivity that is as good as or better than these current models [79].

In our hands, the zebrafish developmental toxicity assay appears to be a reasonably sensitive assay for screening for developmental anomalies of our internal Johnson & Johnson library compounds. If rat data alone are used for comparison, and rabbit and/or mouse data ignored from the mammalian testing, 86% sensitivity is calculated, whereas if rabbit data alone are taken into account for these calculations, no false-negative results are obtained (100% sensitivity) [81]. But since it is unclear whether rat, rabbit, or multiple species represent the best predictive value for human teratogenesis, a combined sensitivity of 75% for compounds of the Janssen portfolio, established by comparing the zebrafish results with the outcome of all available mammalian *in vivo* studies, may be more

Table 11.1 Example of morphological characteristics evaluated as measurements for the teratogenic potency of a test compound at designed time points: comparison between three test setups.

Brannen *et al.* [80]		Van den Bulck *et al.* [81]	Selderslaghs *et al.* [82]
Morphological endpoint	**Method of assessment**	**Morphological endpoint: teratogenic endpoints**	**Morphological endpoint: zebrafish embryo**
Body length[a)]	Measured in millimeters	Developmental retardation/arrest	Egg coagulation
Body shape	Morphological score and description	Body shape abnormal	Somites
Viability	Percent incidence in treatment group	Eye abnormalities	Tail detachment
Head–trunk angle[a)]	Measured in degrees	Ear otolith abnormalities	Otolith
Otic vesicle length[a)]	Estimated distance between eye and otic vesicle	Brain/spinal cord abnormalities	Eyes
Somite number[a)]	Somite pairs counted	Heart morphology	Heartbeat
Somite morphology	Morphological score and description	Anemia	Blood circulation
Notochord morphology	Morphological score and description	Angiogenesis	*Morphological endpoint: zebrafish larvae*
Tail morphology	Morphological score and description	Pigment abnormalities	Skeletal abnormalities
Fin morphology	Morphological score and description	Jaw abnormalities	Sidewise position
Heart morphology	Morphological score and description	Dorsal/caudal fin abnormalities	Active swimming (upon stimulation)
Cardiovascular function	Normal/abnormal and description	Pectoral fin abnormalities	
Facial structure morphology	Morphological score and description	*Morphological endpoint: embryotoxic endpoints*	
Brain morphology	Morphological score and description	Reduced motility/touch response	
Jaw and pharyngeal arch morphology	Normal/abnormal and description	Hyperactivity	
Motility	Normal/abnormal and description	Unhatched	
Pigmentation	Normal/abnormal and description	Necrosis	
Swim bladder[a)]	Normal/abnormal and description	Necrosis and cranial edema	
Stomach[a)]	Normal/abnormal and description	Necrosis and enlarged liver	
Intestine	Normal/abnormal and description	Heart function abnormal	
Liver[a)]	Normal/abnormal and description	Pericardial edema	
Yolk	Normal/abnormal and description	Circulation slow or absent	
Heart rate[a)]	Counted in beats per minute	Skin abnormalities Gut abnormalities Necrosis, nephric cysts Yolk abnormalities, yolk sac edema	

a) Evaluated for possible use as teratogenic endpoint but not retained in the final scoring system.

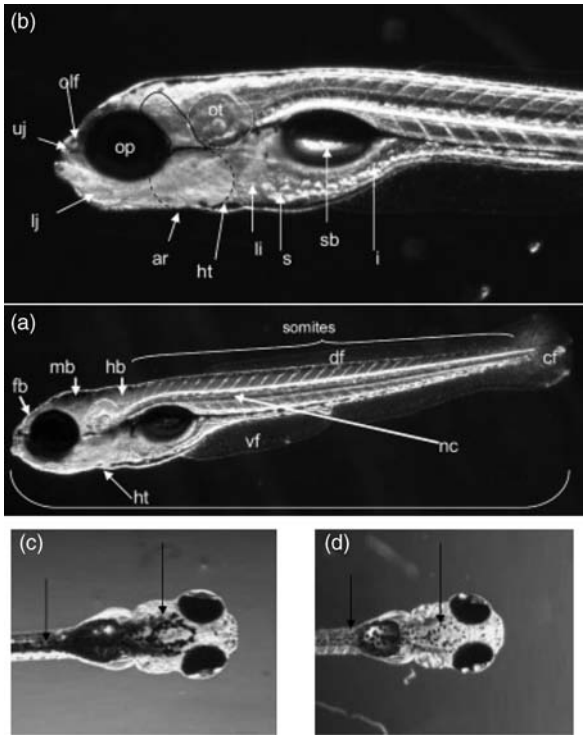


Figure 11.4 Morphology of the 5 dpf zebrafish larva. (a) Lateral view of the 5 dpf zebrafish. Bracket (bottom) defines the head–tail region measured for larva length. fb: forebrain; mb: midbrain; hb: hindbrain; df: dorsal fin; cf: caudal fin; vf: ventral fin; nc: notochord. (b) Close-up of anterior region. olf: olfactory region; uj: upper jaw; lj: lower jaw; ar: pharyngeal arches; ht: heart; li: liver; s: stomach; sb: swim bladder; i: intestine. (c) Larva with normal pigmentation level and melanophore pattern (arrows). (d) Larva with reduced pigmentation but normal melanophore pattern (arrows). (From Ref. [80].)

applicable. This outcome is comparable to the sensitivity recently reported in previous studies using this model; however, to use this screen as a valuable and reliable tool in drug selection and development, adequate specificity is a prerequisite, which is not achieved for the Janssen set of compounds based on the evaluation of the current set of zebrafish endpoints, even when corrected for bioanalysis (43, 50, and 38% specificity for all mammalian data, rat-only data, and rabbit-only data, respectively). The reason for the lower specificity of the test system when compared with literature data remains unclear. The chemical class of compounds, uptake and possible metabolism, and the mode of action of the drugs are important factors to take into account for the interpretation of these results. The large set of endpoints scored may also account for the discrepancy. Historically, both the number of endpoints assessed and the length of the exposure and assessment period varied considerably, although there is a tendency for

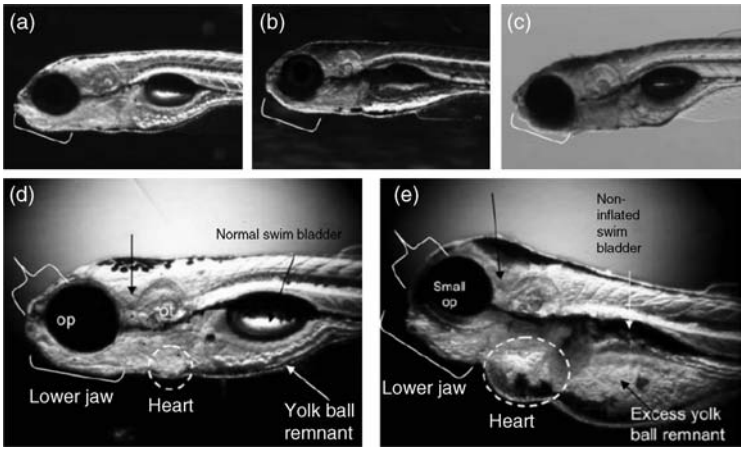


Figure 11.5 Examples of craniofacial and visceral dysmorphology with assigned morphological scores. Bracket defines the lower jaw in (a)–(e). (a) Anterior lateral view of a control 5 dpf larva with normal morphology (scores = 5). (b) Larva treated with BMS-4 (endothelin receptor antagonist) with subtle reduction in the size of the lower jaw (scores = 4). (c) Larva treated with BMS-4 presenting definitive deficiency of the lower jaw (scores = 3). (d) Close-up of craniofacial region of a normal control larva with normal spacing (black arrow) found between the optic (op) and otic (ot) vesicles; short bracket: forebrain/olfactory region/upper jaw (scores of all structures = 5). (e) Close-up of the craniofacial region of a larva exposed to a CB-1 antagonist, BMS-7, presenting marked or severe dysmorphology of various structures. Short bracket: significantly reduced forebrain, olfactory region, and upper jaw. Irregularly shaped and shortened lower jaw (lj) (upper and lower jaw scores = 1); heart (circled) with swollen pericardial sac and irregularly shaped chambers (scores = 1); narrowed spacing (black arrow) between optic and otic vesicles; and reduced optic vesicle (op) size (facial scores = 2). (From Ref. [80].)

less endpoints and a longer exposure period starting shortly after fertilization. For example, the new OECD guidelines for the fish embryo toxicity (FET) test drafted in 2006 [87] include observations at 24 and 48 hpf and only four apical endpoints: embryo coagulation, irregular somite formation, nondetachment of the tail, and lack of heartbeat. In a similar fashion, Selderslaghs *et al.* [82] also assessed these endpoints prehatch (up to 72 hpf), but also scored the otic vesicle and otoliths, eyes, and blood circulation during this period, and then evaluated larval motility, skeletal deformities, and body position at 144 hpf (Table 11.1). In our current assay, we opted for daily evaluation of vitality and developmental assessment with binomial scoring of a series of endpoints at 96 hpf (Table 11.1). In contrast, a quantitative method was described by Brannen *et al.* [80], in which similar anatomical structures as in our assay were assessed at 120 hpf, together with cardiovascular function and motility. Larval body length, head–trunk angle, and eye–otic vesicle distance were measured, as well as scoring of the severity of each phenotype (Table 11.1). Overall, that study protocol seemed to be more comparable to the OECD 212 test guideline [88], which additionally included assessments on behavior, hatching success, yolk sac absorption, and survival

as well as length and weight measurements. Future collaborations should work on defining the exact endpoints and exposure times toward the harmonization of a global zebrafish teratogenicity protocol before the assay can be used in safety assessment.

11.3.5 Future Directions: ADME Studies and Future Explorative Research

11.3.5.1 Absorption and Distribution

Defining the effects of ADME on drugs and their metabolites in the zebrafish will be vital as the organism becomes more widely used in toxicology. Although it has been shown that zebrafish toxicological assays can attain a good level of predictivity, false negatives and false positives have been found to compromise the sensitivity and specificity of the assays used. These studies, while sporadic in nature, underscored a neglected understanding of the absorption, distribution, metabolism, and excretion (ADME) profile of drugs between zebrafish and human or other mammalian models. The ADME profile of a drug in turn is affected by factors such as the route of administration, physicochemical properties of the drug, and physiology of the fish. To fully realize the potential of zebrafish as a drug toxicological model, the knowledge gap in ADME needs to be addressed. In using zebrafish embryo as a high-throughput screening system, the most common route of drug administration is by immersion of embryo into a medium that contains the solubilized drug. Researchers should bear in mind the fact that the immersion technique for dosing of drugs to zebrafish is analogous to depot injection in the mammalian model as absorption and continuous exposure of the biological system to the drug are present in both cases. This information is important for selecting a suitable mathematical model for pharmacokinetic study in zebrafish. The drug is then absorbed by the embryo across the chorion enveloping the embryo (if the drug fails to absorb through the chorion, then the chorion can be removed chemically or manually). In the larvae or adult fish, drug absorption occurs through the skin, gill, and gut, out of which dermal uptake may contribute to 50% of the total uptake of compounds that are hydrophobic [89]. Realizing the importance of ascertaining the extent to which a drug is absorbed by zebrafish embryo to circumvent erroneous pharmacokinetic–pharmacodynamic correlation, researchers would benefit if there exists a set of rules comprising cutoff molecular weight, log *P*, or polar surface area, such as Lipinski's rule of five for the determination of druglikeness of compounds [90], to help predict whether a compound can be absorbed into the zebrafish embryo. Evaluation of the uptake of a wide range of compounds would be required in order to derive such rules. As a recommendation, *in vivo* drug absorption studies might have to be conducted prior to toxicological testing using zebrafish. Alternatively, to increase throughput, drug absorption studies could be conducted only for potential hits that yielded negative results in the toxicological assay. The extent of the distribution of a drug to a particular organ or tissue is of importance, especially when studying specific organ toxicity. Distribution of drug is

affected by blood flow rate, binding to plasma/tissue proteins, permeability across membrane barriers, and interaction with transporters. It has been shown that the total concentration of plasma protein as well as qualitative properties of individual protein classes differs between fish and mammals [89]. This difference would certainly impact the drug distribution between the plasma and tissue compartments in zebrafish compared with mammals. Equilibrium dialysis methods may be used to investigate the plasma protein binding of drug in zebrafish. However, one must be mindful of the possible challenge in obtaining sufficient volume of zebrafish blood to conduct such experiments. Investigation of drug distribution in zebrafish using a panel of model drugs with diversified physicochemical properties is a necessary step forward in future studies. Drug absorption from aqueous media by the larval fish is highly dependent on pH, so care must be taken to ensure that the media for chemical screens are suitably buffered. The physicochemical properties of most small molecules can be used to predict absorption, and these properties correlate with the usual attributes of successful drugs. The objective measurement of drug absorption as well as drug distribution can be readily assessed using fluorescence or radioactive assays [91,92].

11.3.5.2 Metabolism

As metabolism is one of the causes of drug-induced toxicities, it is imperative to characterize biotransformation capabilities of zebrafish and compare it against human or a current preclinical model such as a rodent. It is also equally important to validate the ability of zebrafish to generate reactive metabolites that have been shown to be responsible for causing drug-induced toxicities [93]. Without this knowledge, researchers will not know, as in most cases when using zebrafish assays, whether the toxicity observed is caused by a parent drug or a reactive metabolite from the drug, and if the latter, whether this reactive metabolite will also be formed in humans. At present, in response to this knowledge gap, the functional activities of phase I enzymes such as CYPs and phase II enzymes such as SULT and UGT are being evaluated [94–98]. Even so, these few studies have only just begun to unravel the similarities and differences of metabolism and bioactivation of drugs between zebrafish, human, and rodent models. In the past decade, drug transporters have been intensely studied as the awareness of their impact on the absorption, distribution, excretion, and toxicity of drugs has increased [99]. However, little work has been done in the field of drug transporter research in zebrafish. A total of 41 ATP binding cassette (ABC) transporters have been identified in the zebrafish genome [100]. In a proof-of-concept experiment, Scholz *et al.* demonstrated that cyclosporin A inhibition of ABC transporter caused accumulation of rhodamine B in zebrafish larvae [101]. This study reinforced the importance of characterizing drug transporters and evaluating their role in zebrafish.

11.3.5.3 Harmonization and Validation

The lack of harmonization of protocols for the various toxicity tests conducted using zebrafish limits the reproducibility of results between experiments and

laboratories. Experimental details (e.g., temperature, pH of media, and choice of solvent/vehicle that could affect ADME of drugs) and toxicological assessment approaches should be standardized in order to validate objectively the accuracy of zebrafish as a predictive toxicological model. The developmental stage of the zebrafish at the time of the experiment initiation is yet another important consideration. If the pharmaceutical scientist is interested in investigating direct toxicity from parent drug, presence of a mature liver may not be crucial. However, a mature metabolizing liver may be critical when a toxic electrophilic metabolite is suspected. Finally, a scaling factor for the *in vivo–in vivo* correlation between the zebrafish and human or rodent models is unavailable. Without a scaling factor, scientists cannot make a meaningful correlation between the toxic dose observed in zebrafish and humans or rodents. It is hoped that with rigorous ADME screening of the zebrafish model using a wide range of drug compounds, scaling factors and chemical rules on drug ADME specific to the zebrafish larvae can be derived. The Zebrafish Consortium comprising pharmaceutical companies such as Bristol-Myers Squibb, Pfizer, and AstraZeneca has started to formulate harmonized teratogenicity assays using the zebrafish model [102] and it is hoped that this will extend to other toxicity assays as well. While not required for screening purposes, thorough validation of the harmonized assays is a prerequisite step in the application of the 3R principle.

According to the ECVAM (European Centre for Validation of Alternative Methods) principles, a new developed assay needs to flow through different stages of performance, applicability, and predictivity before the assay can be accepted as a valid test [103]. The first module, the *test (model) definition*, includes the establishment of a test protocol and the definition of the endpoint predicted to set up a test model, the zebrafish assays. The models need to be trained with a training set of compounds, a selected set of reference compounds, for which the mechanism of action is known. A training set of compounds allows testing the provisional domain of applicability on different pharmaceutical in-house chemistries. Both sets of compounds are too small to validate the assay (numbers are smaller than or equal to around 50, see below) but will provide valuable predictivity indices to start with and insights into important mechanisms of drug-induced toxicity in the zebrafish. This should then result in the definition of a prediction model. *Within-laboratory variability*, *transferability*, and *between-laboratory variability* are the next steps that need to be performed. *Predictive capacity* is a very important stage, as the data set will need to be enlarged to further explore the predictivity indices of the zebrafish model. As a training set consists of only 50 compounds, specificity and sensitivity indices are not representative for a scenario in which 500 compounds would be tested. It could be calculated that for an expected specificity of 90% (the zebrafish assays would detect 10% false positives) a data set of almost 500 compounds evenly distributed between toxic and nontoxic compounds would be needed to obtain a lower confidence limit of 80%, meaning that while repeating the experiment, the actual specificity will vary between 90 and 80% [104]. This calculation only demonstrates the challenges associated with introducing new predictive tests into the

area of toxicology. In addition, the *application domain* will also be a very important follow-up project, as the applicability of the test model will need to be enlarged to the whole pharmaceutical compound libraries. Finally, *performance standards* can be explored, when there is interest from regulatory authorities for this specific test. This includes not only incorporation into an ECVAM validation, but also interest from FDA and EMEA in accepting alternative tests for studying drug-induced toxicity, in addition to or instead of the traditionally performed animal testing for assessing embryotoxicity of chemicals in the REACH program, for example.

11.3.5.4 Future Explorative Research

The zebrafish occupies an important niche between more traditional representative animal models and tractable lower organisms or *in vitro* systems (Figure 11.1). Although often viewed as a cheap alternative to rodents, the zebrafish offers comprehensive vertebrate pathway and cellular context on a scale hitherto feasible only in cell culture. If we are to maximize the utility of the zebrafish in discovery mode, we will require a much more nuanced understanding of the parallels between the zebrafish and higher organisms in health and in disease. It will be essential to develop more global screening assays, focused not just on previously described toxicities, but also on detecting the unforeseen. An unbiased phenoclustering strategy combining multichannel organ-specific reporters, functional genomics, and automated image analysis may be feasible in the near future, and this would offer significant advantages over current low-throughput toxicity studies in inbred rodent strains. The resultant prospect of systematic exploration of gene–drug, drug–drug, and drug–environment interactions is highly attractive. The ability to predict specific toxicities and to define the modifying effects of genome, microbiome, epigenome, and environment is a major milestone in realizing the vision of predictive toxicology. As these goals are attained, the role of the zebrafish as a tool for annotation of chemical libraries is likely to expand, moving its use to earlier in the drug discovery pipeline. It is conceivable that with parallel efforts in zebrafish disease modeling drug discovery and predictive toxicology may eventually occur in parallel, thus enabling direct optimization of therapeutic efficacy and toxicity for each disease context. A robust zebrafish screening platform may be one of the few approaches able to deliver on these promises, but it will require large investments in the coming years to fill the remaining gaps and unravel the true predictive power of this small fish.

References

1 Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.*, **3**, 711–715.
2 Stevens, J. and Baker, T. (2009) The future of drug safety testing: expanding the view and narrowing the focus. *Drug Discov. Today*, **14**, 162–167.
3 Olsen, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W.,

Dorate, M., Van Deun, K., Smith, P., Berger, B., and Heller, A. (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.*, **32**, 56–67.

4 Hill, A.J., Teraoka, H., Heideman, W., and Peterson, R.E. (2005) Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicol. Sci.*, **86**, 6–19.

5 Chakraborty, C., Hsu, C.H., Wen, Z.H., Lin, C.S., and Agoramoorthy, G. (2009) Zebrafish: a complete animal model for *in vivo* drug discovery and drug development. *Curr. Drug Metab.*, **10**, 116–124.

6 Gehrig, J., Reischl, M., and Kalmar, E. (2009) Automated high-throughput mapping of promoter–enhancer interactions in zebrafish embryos. *Nat. Methods*, **6**, 911–916.

7 Pardo-Martin, C., Chang, T.Y., and Koo, B.K. (2010) High-throughput *in vivo* vertebrate screening. *Nat. Methods*, **7**, 634–636.

8 Sukardi, H., Chng, H.T., Chan, E.C., Gong, Z., and Lam, S.H. (2011) Zebrafish for drug toxicity screening: bridging the *in vitro* cell-based models and *in vivo* mammalian models. *Expert Opin. Drug Metab. Toxicol.*, **7**, 579–589.

9 Augustine-Rauch, K., Zhang, C.X., and Panzica-Kelly, J.M. (2010) *In vitro* developmental toxicology assays: a review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays. *Birth Defects Res. C Embryo Today*, **90**, 87–98.

10 Strahle, U. and Grabher, C. (2010) The zebrafish embryo as a model for assessing off-target drug effects. *Dis. Models Mech.*, **3**, 689–692.

11 Mittelstadt, S.W., Hemenway, C.L., Craig, M.P., and Hover, J.R. (2008) Evaluation of zebrafish embryos as a model for assessing inhibition of hERG. *J. Pharmacol. Toxicol. Methods*, **57**, 100–105.

12 Rubenstein, A.L. (2006) Zebrafish assays for drug toxicity screening. *Expert Opin. Drug Metab. Toxicol.*, **2**, 231–240.

13 Spitsbergen, J.M. and Kent, M.L. (2003) The state of the art of the zebrafish model for toxicology and toxicologic pathology research – advantages and current limitations. *Toxicol. Pathol.*, **31** (Suppl.), 62–87.

14 Lammer, E., Carr, G.J., and Wendler, K. (2009) Is the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test? *Comp. Biochem. Physiol. C Toxicol. Pharmacol.*, **149**, 196–209.

15 Roden, D.M. (1998) Mechanisms and management of proarrhythmia. *Am. J. Cardiol.*, **82**, 491–557.

16 Keating, M.T. and Sanguinetti, M.C. (2001) Molecular and cellular mechanisms of cardiac arrhythmias. *Cell*, **104**, 569–580.

17 Camm, A.J., Janse, M.J., Roden, D.M., Rosen, M.R., Cinca, J., and Cobbe, S.M. (2000) Congenital and acquired long QT syndrome. *Eur. Heart J.*, **21**, 1232–3781.

18 Roy, M., Dumaine, R., and Brown, A.M. (1996) HERG, a primary human ventricular target of the nonsedating antihistamine terfenadine. *Circulation*, **94**, 817–823.

19 Cavero, I., Mestre, M., Guillon, J.M., and Crumb, W. (2000) Drugs that prolong QT interval as an unwanted effect: assessing their likelihood of inducing hazardous cardiac dysrhythmias. *Expert Opin. Pharmacother.*, **1**, 947–973.

20 Mitcheson, J.S., Chen, J., Lin, M., Culberson, C., and Sanguinetti, M.C. (2000) A structural basis for drug-induced long QT syndrome. *Proc. Natl. Acad. Sci. USA*, **97**, 12329–12333.

21 Yang, T., Snyders, D., and Roden, D.M. (2001) Drug block of I_{Kr}: model systems and relevance to human arrhythmias. *J. Cardiovasc. Pharmacol.*, **38**, 737–744.

22 Roden, D.M. (2001) Pharmacogenetics and drug-induced arrhythmias. *Cardiovasc. Res.*, **50**, 224–231.

23 Abbott, G.W., Sesti, F., Splawski, I., Buck, M.E., and Lehmann, M.H. (1999) MiRP1 forms I_{Kr} potassium channels with HERG and is associated with cardiac arrhythmia. *Cell*, **97**, 175–187.

24 Furutani, M., Trudeau, M.C., Hagiwara, N., Seki, A., and Gong, Q. (1999) Novel mechanism associated with an inherited cardiac arrhythmia: defective protein

trafficking by the mutant HERG (G601S) potassium channel. *Circulation*, **99**, 2290–2294.
25 Martin, R.L., McDermott, J.S., Salmen, H.J., Palmatier, J., Cox, B.F., and Gintant, G.A. (2004) The utility of hERG and repolarization assays in evaluating delayed cardiac repolarization: influence of multi-channel block. *J. Cardiovasc. Pharmacol.*, **43**, 369–379.
26 Thisse, C. and Zon, L.I. (2002) Organogenesis – heart and blood formation from the zebrafish point of view. *Science*, **295**, 457–462.
27 Langheinrich, U. (2003) Zebrafish: a new model on the pharmaceutical catwalk. *Bioessays*, **25**, 904–912.
28 Berghmans, S. (2006) Zebrafish: validation of a pro-arrhythmia *in vivo* model (abstract for oral communication). IIR's 4th Annual QT Prolongation and Safety Pharmacology Conference: Predicative Safety Assessment in Non-Clinical and Clinical Studies, February 28–March 3, Amsterdam, The Netherlands.
29 Milan, D.J., Peterson, T.A., Ruskin, J.N., Peterson, R.T., and MacRae, C.A. (2003) Drugs that induce repolarization abnormalities cause bradycardia in zebrafish. *Circulation*, **107**, 1355–1358.
30 Berghmans, S., Butler, B., Goldsmith, P., Waldron, G., Gardner, I., and Golder, Z. (2008 Zebrafish based assays for the assessment of cardiac, visual and gut function – potential safety screens for early drug discovery. *J. Pharmacol. Toxicol. Methods*, **58**, 59–68.
31 Burns, C.G., Milan, D.J., and Grande, E.J. (2005) High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nat. Chem. Biol.*, **1**, 263–264.
32 MacRae, C.A. (2010) Cardiac arrhythmia: *in vivo* screening in the zebrafish to overcome complexity in drug discovery. *Expert Opin. Drug Discov.*, **5**, 619–632.
33 Shin, J.T., Pomerantsev, E.V., Mably, J.D., and MacRae, C.A. (2010) High-resolution cardiovascular function confirms functional orthology of myocardial contractility pathways in zebrafish. *Physiol. Genomics*, **42**, 300–309.
34 Tropepe, V. and Sive, H.L. (2003) Can zebrafish be used as a model to study the neurodevelopmental causes of autism? *Genes Brain Behav.*, **2**, 268–281.
35 Rink, E. and Wulliman, M.F. (2004) Connections of the ventral telencephalon (subpallium) in the zebrafish (*Danio rerio*). *Brain Res.*, **1011**, 206–220.
36 Wullimann, M.F. and Mueller, T. (2004) Teleostean and mammalian forebrains contrasted: evidence from genes to behavior. *J. Comp. Neurol.*, **478**, 427–428.
37 Zhdanova, I.V., Wang, S.Y., Leclair, O.U., and Danilova, N.P. (2001) Melatonin promotes sleep-like state in zebrafish. *Brain Res.*, **903**, 263–268.
38 Leimer, U., Lun, K., Romig, H., Walter, J., Grunber, J., Brand, M., and Haass, C. (1999) Zebrafish (*Danio rerio*) presenilin promotes aberrant amyloid peptide production and requires a critical aspartate residue for its function in amyloidogenesis. *Biochemistry*, **38**, 13602–13609.
39 Monnot, M.J., Babin, P.J., Andre, M., Laforest, L., Ballagny, C., and Akimenko, M.A. (1999) Epidermal expression of apolipoprotein E gene during fin and scale development and fin regeneration in zebrafish. *Dev. Dyn.*, **214**, 207–215.
40 Behra, M., Etard, C., Cousin, X., and Strahle, U. (2004) The use of zebrafish mutants to identify secondary target effects of acetylcholine esterase inhibitors. *Toxicol. Sci.*, **77**, 325–333.
41 Karlovich, C.A., John, R.M., Ramirez, L., Stainier, D.Y., and Myers, R.M. (1998) Characterization of the Huntington's disease (HD) gene homologue in the zebrafish *Danio rerio*. *Gene*, **217**, 117–125.
42 Rihel, J., Prober, D.A., Arvanites, A., Lam, K., and Zimmerman, S. (2010) Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science*, **327**, 348–351.
43 Kokel, D., Bryan, J., Laggner, C., White, R., and Cheung, C.Y. (2010) Rapid behavior-based identification of neuroactive small molecules in the zebrafish. *Nat. Chem. Biol.*, **6**, 231–237.
44 Winter, M.J., Redfern, W.S., Hayfield, A. J., Owen, S.F., Valentin, J.P., and

Hutchinson, T.H. (2008) Validation of a larval zebrafish locomotor assay for assessing the seizure liability of early-stage development drug. *J. Pharmacol. Toxicol. Methods*, **57**, 176–187.

45 Barros, T.P., Alderton, W.K., Reynolds, H.M., Roach, A.G., and Berghmans, S. (2008) Zebrafish: an emerging technology for *in vivo* pharmacological assessment to identify potential safety liabilities in early drug discovery. *Br. J. Pharmacol.*, **154**, 1400–1413.

46 Ostapowicz, G., Fontana, R.J., Schiodt, F. V., Larson, A., Davern, T.J., and Han, S.H. (2002) Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann. Intern. Med.*, **137**, 947–954.

47 Reuben, A., Koch, D.G., and Lee, W.M. (2010) Drug-induced acute liver failure: results of a U.S. multicenter, prospective study. *Hepatology*, **52**, 2065–2076.

48 Navarro, V.J. and Senior, J.R. (2006) Drug-related hepatotoxicity. *N. Engl. J. Med.*, **354**, 731–739.

49 Abboud, G. and Kaplowitz, N. (2007) Drug-induced liver injury. *Drug Saf.*, **30**, 277–294.

50 O'Brien, P.J., Slaughter, M.R., Biagini, C., Diaz, D., Gao, B., and Irwin, W. (2003) Predicting drug-induced human hepatotoxicity with *in vitro* cytotoxicity assays. Proceedings Tox'03, London, UK.

51 Amacher, D.E. (2011) Strategies for the early detection of drug induced hepatic steatosis in preclinical drug safety evaluation studies. *Toxicology*, **279**, 10–18.

52 Xu, J.J., Henstock, P.V., Dunn, M.C., Smith, A.R., Chabot, J.R., and De Graaf, D. (2008) Cellular imaging predictions of clinical drug induced liver injury. *Toxicol. Sci.*, **105**, 97–105.

53 Cosgrove, B.D., King, B.M., Hasan, M.A., Alexopoulos, L.G., Farazi, P.A., and Hendriks, B.S. (2009) Synergistic drug–cytokine induction of hepatocellular death as an *in vitro* approach for the study of inflammation-associated idiosyncratic drug hepatotoxicity. *Toxicol. Appl. Pharmacol.*, **237**, 317–330.

54 Cosgrove, B.D., Alexopoulos, L.G., Hang, T.-C., Hendriks, B.S., Sorger, P.K., and Griffith, L.G. (2010) Cytokine-associated drug toxicity in human hepatocytes is associated with signaling network dysregulation. *Mol. BioSyst.*, **6**, 1195–1206.

55 Roth, R.A., Harkema, J.R., Pestka, J.P., and Ganey, P.E. (1997) Is exposure to bacterial endotoxin a determinant of susceptibility to intoxication from xenobiotic agents? *Toxicol. Appl. Pharmacol.*, **147**, 300–311.

56 Lee, Y.H., Chung, M.C., Lin, Q., and Boelsterli, U.A. (2008) Troglitazone induced hepatic mitochondrial proteome expression dynamics in heterozygous $Sod2^{+/-}$ mice: two-stage oxidative injury. *Toxicol. Appl. Pharmacol.*, **231**, 43–51.

57 Kashimshetty, R., Desai, V.G., Kale, V.M., Lee, T., Moland, C.L., and Branham, W.S. (2009) Underlying mitochondrial dysfunction triggers flutamide-induced oxidative liver injury in a mouse model of idiosyncratic drug toxicity. *Toxicol. Appl. Pharmacol.*, **238**, 150–159.

58 Chu, J. and Sadler, K.C. (2009) New school in liver development: lessons from zebrafish. *Hepatology*, **50**, 1656–1663.

59 Pack, M., Solnica-Krezel, L., Malicki, J., Neuhauss, S.C., Schier, A.F., Stemple, D. L., Driever, W., and Fishman, M.C. (1996) Mutations affecting development of zebrafish digestive organs. *Development*, **123**, 321–328.

60 Lorent, K., Yeo, S.Y., Oda, T., Chandrasekharappa, S., Chitnis, A., Matthews, R.P., and Pack, M. (2004) Inhibition of Jagged-mediated Notch signaling disrupts zebrafish biliary development and generates multi-organ defects compatible with an Alagille syndrome phenocopy. *Development*, **131**, 5753–5766.

61 Sadler, K.C., Amsterdam, A., Soroka, C., Boyer, J., and Hopkins, N. (2005) A genetic screen in zebrafish identifies the mutants vps18, nf2 and foie gras as models of liver disease. *Development*, **132**, 3561–3572.

62 Lam, S.H. and Gong, Z. (2006) Modeling liver cancer using zebrafish: a comparative oncogenomics approach. *Cell Cycle*, **5**, 573–577.

63 Braunbeck, T., Gorge, G., Storch, V., and Nagel, R. (1990) Hepatic steatosis in zebrafish (*Brachydanio rerio*) induced by long-term exposure to gamma-hexachlorocyclohexane. *Ecotoxicol. Environ. Saf.*, **19**, 355–374.

64 Amali, A.A., Rekha, R.D., Lin, C.J., Wang, W.L., Gong, H.Y., Her, G.M., and Wu, J.L. (2006) Thioacetamide induced liver damage in zebrafish embryo as a disease model for steatohepatitis. *J. Biomed. Sci.*, **13**, 225–232.

65 Passeri, M.J., Cinaroglu, A., Gao, C., and Sadler, K.C. (2009) Hepatic steatosis in response to acute alcohol exposure in zebrafish requires sterol regulatory element binding protein activation. *Hepatology*, **49**, 443–452.

66 McGrath, P. and Li, C.Q. (2008) Zebrafish: a predictive model for assessing drug-induced toxicity. *Drug Discov. Today*, **13**, 394–401.

67 Hill, A.J. (2011) Hepatotoxicity testing in larval zebrafish, in *Zebrafish: Methods for Assessing Drug Safety and Toxicity* (ed. P. McGrath), Wiley-Blackwell, West Sussex, UK.

68 Hill, A., Mesens, N., Steemans, M., Xu, J.J., and Aleo, M.D. (2012) Comparisons between *in vitro* whole cell imaging and *in vivo* zebrafish-based approaches for identifying potential human hepatotoxicants earlier in pharmaceutical development. *Drug Metab. Rev.*, **44**, 127–140.

69 Zhang, C., Willett, C., and Fremgen, T. (2001) *Zebrafish: An Animal Model for Toxicological Studies*, John Wiley & Sons, Inc., New York.

70 He, J.H., Guo, S.Y., Zhu, F., Zhu, J.J., Chen, Y.X., Huang, C.J., Gao, J.M., Dong, Q.X., Xuan, Y.X., and Li, C.Q. (2013) A zebrafish phenotypic assay for assessing drug-induced hepatotoxicity. *J. Pharmacol. Toxicol. Methods*, **67** (1), 25–32.

71 Driessen, M., Kienhuis, A.S., Pennings, J.A.L., Pronk, T.E., van de Brandhof, E.J., Roodbergen, M., Spaink, H.P., van de Water, B., and van der Ven, L.T.M. (2013) Exploring the zebrafish embryo as an alternative model for the evaluation of liver toxicity by histopathology and expression profiling. *Arch. Toxicol.*, **87**, 807–823.

72 ICH (2009) M3(R2): Guidance on Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. ICH Harmonized Tripartite Guideline endorsed by ICH Steering Committee, Washington, DC.

73 ICH (1993) Guideline on Detection of Toxicity of Reproduction for Medicinal Products. ICH Harmonized Tripartite Guideline endorsed by ICH Steering Committee, Washington, DC.

74 Piersma, A.H. (2004) Validation of alternative methods for developmental toxicity testing. *Toxicol. Lett.*, **149**, 147–153.

75 Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995) Stages of embryonic development of the zebrafish. *Dev. Dyn.*, **203** (3), 253–310.

76 Peterson, R.E., Theobald, H.M., and Kimmel, G.L. (1993) Developmental and reproductive toxicity of dioxins and related compounds: cross-species comparisons. *Crit. Rev. Toxicol.*, **23** (3), 283–335.

77 Ito, T., Ando, H., Suzuki, T., Ogura, T., Hotta, K., Imamura, Y., Yamaguchi, Y., and Handa, H. (2010) Identification of a primary target of thalidomide teratogenicity. *Science*, **327**, 1345–1350.

78 Panzica-Kelly, J.M., Zhang, C.X., and Danberry, T.L. (2010) Morphological score assignment guidelines for the dechorionated zebrafish teratogenicity assay. *Birth Defects Res. B Dev. Reprod. Toxicol.*, **89**, 382–395.

79 Chapin, R., Augustine-Rauch, K., and Beyer, B. (2008) State of the art in developmental toxicity screening methods and a way forward: a meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. *Birth Defects Res. B Dev. Reprod. Toxicol.*, **83**, 446–456.

80 Brannen, K.C., Panzica-Kelly, J.M., Danberry, T.L., and Augustine-Rauch, K.A. (2010) Development of a zebrafish embryo teratogenicity assay and

quantitative prediction model. *Birth Defects Res. B Dev. Reprod. Toxicol.*, **89**, 66–77.

81 Van den Bulck, K., Hill, A., Mesens, N., Diekman, H., De Schaepdrijver, L., and Lammens, L. (2011) Zebrafish developmental toxicity assay: a fishy solution to reproductive toxicity screening, or just a red herring? *Reprod. Toxicol.*, **32**, 213–219.

82 Selderslaghs, I., Van Rompay, A.R., De Coen, W., and Witters, H.E. (2009) Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo. *Reprod. Toxicol.*, **28**, 308–320.

83 Peterson, R.T., Link, B.A., Dowling, J.E., and Schreiber, S.L. (2000) Small molecule developmental screens reveal the logic and timing of vertebrate development. *Proc. Natl. Acad. Sci. USA*, **97** (24), 12965–12969.

84 Busquet, F., Nagel, R., von Landenberg, F., Mueller, S.O., Huebler, N., and Broschard, T.H. (2008) Development of a new screening assay to identify proteratogenic substances using zebrafish *Danio rerio* embryo combined with an exogenous mammalian metabolic activation system (mDarT). *Toxicol. Sci.*, **104** (1), 177–188.

85 Nagel, R. and Dar, T. (2002) The embryo test with the zebrafish *Danio rerio* – a general model in ecotoxicology and toxicology. *ALTEX*, **19** (Suppl. 1), 38–48.

86 Eimon, P.M. and Rubinstein, A.L. (2009) The use of *in vivo* zebrafish assays in drug toxicity screening. *Expert Opin. Drug Metab. Toxicol.*, **5**, 393–401.

87 OECD (2006) Draft Test Guideline on "Fish Embryo Toxicity Test". Organisation for Economic Cooperation and Development, Paris, France. Available at http://www.oecd.org/dataoecd/39/59/36817070.pdf (last accessed September 18, 2013).

88 OECD (1998) OECD Guidelines for the Testing of Chemicals. No. 212: Fish, Short-Term Toxicity Test on Embryo and Sac-Fry Stages. Organisation for Economic Cooperation and Development, Paris, France.

89 Kleinow, K.M., Nicols, J.W., and Hayton, W.L. (2008) Toxicokinetics in fishes, in *The Toxicology of Fishes* (eds R.T. Di Giulio and D.E. Hinton), CRC Press, pp. 55–152.

90 Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.*, **46**, 3–26.

91 Farber, S.A., Pack, M., Ho, S.Y., Johnson, I.D., and Wagner, D.S. (2001) Genetic analysis of digestive physiology using fluorescent phospholipid reporters. *Science*, **292**, 1385–1388.

92 Koide, K., Song, F., de Groh, E.D., Garner, A.L., and Mitchell, V.D. (2008) Scalable and concise synthesis of dichlorofluorescein derivatives displaying tissue permeation in live zebrafish embryos. *ChemBioChem*, **9**, 214–218.

93 Williams, D.P., Kitteringham, N.R., Naisbitt, D.J. *et al.* (2002) Are chemically reactive metabolites responsible for adverse reactions to drugs? *Curr. Drug Metab.*, **3**, 351–366.

94 Alderton, W., Berghmans, S., and Butler, P. (2010) Accumulation and metabolism of drugs and CYP probe substrates in zebrafish larvae. *Xenobiotica*, **40**, 547–557.

95 Jones, H.S., Panter, G.H., Hutchinson, T.H., and Chipman, J.K. (2010) Oxidative and conjugative xenobiotic metabolism in zebrafish larvae *in vivo*. *Zebrafish*, **7**, 23–30.

96 Scornaienchi, M.L., Thornton, C., and Willett, K.L. (2010) Functional differences in the cytochrome P450 1 family enzymes from zebrafish (*Danio rerio*) using heterologously expressed proteins. *Arch. Biochem. Biophys.*, **502**, 17–22.

97 Kurogi, K., Dillon, J., and Nasser, A. (2010) Sulfation of drug compounds by the zebrafish cytosolic sulfotransferases (SULTs). *Drug Metab. Lett.*, **4**, 62–68.

98 Liu, T.A., Bhuiyan, S., and Liu, M.Y. (2010) Zebrafish as a model for the study of the phase II cytosolic sulfotransferases. *Curr. Drug Metab.*, **11**, 538–546.

99 Endres, C.J., Hsiao, P., Chung, F.S., and Unadkat, J.D. (2006) The role of transporters in drug interactions. *Eur. J. Pharm. Sci.*, **27**, 501–517.
100 Dean, M. and Annilo, T. (2005) Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu. Rev. Genomics Hum. Genet.*, **6**, 123–142.
101 Scholz, S., Fischer, S., and Gundel, U. (2008) The zebrafish embryo model in environmental risk assessment – applications beyond acute toxicity testing. *Environ. Sci. Pollut. Res. Int.*, **15**, 394–404.
102 Augustine-Rauch, K. (2013) Generation and preliminary assessment of a zebrafish teratogenicity assay: an overview of how BMS generated the assay. Bristol-Myers Squibb. Available at http://www.hesiglobal.org/files/public/Committees/DART/Meetings%20and%20Workshops/AlternativeAssayWkshp_Augustine.pdf (last accessed October 10, 2013).
103 Hartung, T., Bremer, S., Casati, S., Coecke, S., Corvi, R., Fortaner, S., Gribaldo, L., Halder, M., Hoffmann, S., Janusch Roi, A., Prieto, P., Sabbioni, E., Scott, L., Worth, A., and Zuang, V. (2004) A modular approach to the ECVAM principles on test validity. *Altern. Lab. Anim.*, **32**, 467–472.
104 Flahault, A., Cadilhac, M., and Thomas, G. (2005) Sample size calculation should be performed for design accuracy in diagnostic test studies. *J. Clin. Epidemiol.*, **58**, 859–862.

12
Predictive Method Development: Challenges for Cosmetics and Genotoxicity as a Case Study[1)]

Gladys Ouédraogo, Fabrice Nesslany, Sophie Simar, Smail Talahari, Doris Lagache, Eric Vercauteren, Lauren Nakab, Astrid Mayoux, Brigitte Faquet, and Nicole Flamand

12.1
Introduction

This chapter addresses different research aspects of alternative/predictive methods in the context of the current legislative arena for cosmetics in Europe. It does not address the regulatory acceptance or the use of these tools for safety assessment. They are described elsewhere in the literature and in the monograph "The SCCS's Notes of Guidance for the Testing of Cosmetic Substances and Their Safety Evaluation" [1] and in Ref. [2].

On July 11, 2013, the European Cosmetics regulation EU Regulation 1223/2009 [3] replaced the Cosmetics directive of 1976 [4]. The ban on animal testing as stated since 2003 remained unchanged in this new regulation. In brief, it stipulates a marketing ban in Europe for cosmetics tested on animals:

- Since September 11, 2004, for finished products.
- Since March 11, 2009, for ingredients tested on short-term toxicity endpoints such as skin irritation, eye irritation, genotoxicity, and acute toxicity.
- Since March 11, 2013, for ingredients tested on long-term toxicity endpoints such as skin sensitization, repeated-dose systemic toxicity, toxicokinetics, and reproductive toxicity.

To address this regulatory context for cosmetics in the European Union (EU), different research initiatives are being led by the European Personal Care Association – Cosmetics Europe – for different endpoints such as skin irritation, skin corrosion, skin sensitization, genotoxicity, and repeated-dose systemic toxicity. For the latter, it is worthwhile highlighting the SEURAT-1 initiative (2011–2015), which is being funded by both the European Commission and

1) The views presented here are those of the authors and do not necessarily reflect the opinion of the company the author is working for.

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

Cosmetics Europe and aims at seeking mode-of-action-based approaches to address repeated-dose systemic toxicity. Cosmetics Europe has been committed for over two decades to the development of alternative methods to animal testing and to fostering their acceptance and use [5].

Also, several tools are being investigated by different stakeholders. The overview given here is far from being comprehensive and will address different types of tools: *in silico* tools and *in vitro* tools (biochemical, 2D cellular assays, and organotypic tools).

Replacing animal studies by alternative methods is a major challenge. It requires great investment in terms of time, budget, and human resources. A key point to consider upfront is problem formulation: Will the tools be used for prioritization for further testing as stand-alone or in combination with other information? Will they contribute to a weight-of-evidence (WoE) process for decision making? Depending on the scope, the level of uncertainty acceptable may vary.

Experience of several years has shown that a one-by-one replacement of *in vivo* assays with alternative/predictive tools is not possible. Moreover, Adler *et al.* concluded that a full replacement of the animal tests used for repeated-dose/reproductive and developmental toxicity testing is not available and the time frame for a full replacement cannot clearly be estimated [6]. The direction is toward building integrated testing strategies (ITS) for different purposes. To build confidence, the methods developed are based on mechanistic considerations, rather than just being descriptive. This is of utmost importance as these mechanistic features may help explain differences observed between *in vivo* and *in vitro* situations and/or interspecies differences. These considerations are in agreement with the recommendations given in the well-known document "Toxicity Testing in the 21st Century" of the National Academy of Sciences (NAS) [7], a vision and a strategy calling for a change of paradigm in toxicity testing. According to this report, hazard and risk assessment can become time and cost effective if we move from descriptive to predictive (mechanistically based) toxicology.

Problem formulation is key when addressing the development of novel tools. Several building blocks need to be considered when engaging in the development of alternative methods:

1) Basic research on the biological models.
   - Need for well-characterized, standardized, and biologically relevant models.
   - Need for a toolbox of models to choose from depending on endpoint/mechanism and intended use.
   - Need for affordable and readily accessible models.
2) A toolbox of methods.
   - To address the physical–chemical diversity of substances.
   - To cover a variety of mechanisms of action: molecular initiating events (MIEs) and key events.

3) Integrated strategies.
   – For building approaches to prioritize chemicals for further testing.
   – To support decision making.

Several research initiatives have addressed or still are addressing the development of alternative/predictive methods for different purposes. A few examples of such initiatives are SEURAT-1 [8], Carcinogenomics [9], Comics [10], ToxCast [11], Tox21 [12], HESS [13], ChemScreen [14], and RISK21 [15].

This chapter will focus on the building blocks for developing predictive approaches. A genotoxicity assay using an organotypic model will be used to illustrate this type of method development.

12.2 The Toolbox of Predictive Methods

12.2.1 *In Silico* Tools

A wealth of data was generated by several companies and organizations worldwide. Information systems have evolved and now are able to handle and process billions of data at once. The next move (which has already started) would be to share data in order to make the best use of what have been done in the past and build the toxicology of the future. Data can be shared within consortia with a confidential and a well-defined legal framework [16]. Building quantitative structure–activity relationships (QSARs) is a way to secure the knowledge generated for decades with different *in vivo* and *in vitro* assays. These models are useful for prioritizing chemicals for further testing. Also, they can guide the choice of the follow-up assay. In addition to safeguarding knowledge and data gap filling, *in silico* models are also hypothesis-generating tools [17,18].

The Organisation for Economic Co-operation and Development (OECD) has provided guidelines for developing transparent and mechanistically based models. So, there is a consensus by all stakeholders (modelers, toxicologists, regulators, and scientists in different fields) on the properties of such models. In addition, the OECD has been working on a QSAR toolbox (the first version was released in 2008 [19]) for Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) [20]. This toolbox hosts not only (Q)SAR models but also databases for forming categories and read-across.

More and more stakeholders are using or recommend the use of the *in silico* tools, provided they comply with the recommendations given in the OECD guidance document, namely,

1) a defined endpoint;
2) an unambiguous algorithm;
3) a defined domain of applicability;

4) appropriate measures of goodness of fit, robustness, and predictivity;
5) a mechanistic interpretation, if possible.

Currently, the vast majority of *in silico* data cannot be used as stand-alone. They are complemented by experimental data.

Typically, *in silico* data can support a weight-of-evidence approach and/or feed into adverse outcome pathway (AOP) approaches. Some *in silico* models cover MIEs or key events [21].

Several *in silico* models are currently available as either freely accessible or commercial tools. They cover different endpoints such as genotoxicity, carcinogenicity, skin sensitization, and phototoxicity. Others address physical–chemical features such as solubility at a given pH, log *P*, log K_{ow}, some ADME (absorption, distribution, metabolism, and excretion) properties such as skin/intestinal permeability, plasma protein binding, and metabolism, or pharmacological targets (interactions with receptor) [22–24]. Careful consideration should be given to their fitness for the purpose they are used for. Their use is restricted to substances with a defined structure.

In silico tools are also used in more complex systems such as physiologically based pharmacokinetic models and virtual tissues to bridge *in vitro* results with *in vivo* effects [25,26].

12.2.2 Biochemical (*In Chemico*) Assays

These assays are usually performed in order to get an insight into some molecular interactions of a test agent with a given (panel of) target(s): receptors or enzymes. They can shed some light on MIEs. *In chemico* data are one piece of information in a WoE approach [27,28].

There are different ways to approach biochemical assays:

- Several targets (receptors and/or enzymes) can be screened at once, when there is nothing specific to focus on or when the objective is to explore widely the ability of a compound to interact with the targets [29]. Usually, this is performed at a given concentration. A follow-up dose–response study is performed for those targets where enzyme activity is reduced by half or where a binding affinity of at least 50% to receptors is achieved. Once the results are generated, the profiles of the test agents can be compared with those from existing substances. This is another way to group chemicals.
- If a specific organ is targeted, the focus can be on a reduced set of targets directly in dose response [30].

In addition, *in chemico* data cannot by themselves inform on a hazard potential of a chemical. Care should be taken to check features such as the bioavailability of the test agent: penetration in the body (by dermal, oral, or inhalation route) and ability to cross biological barriers (e.g., blood–brain barrier

and intestinal permeability). Toxicokinetics can help derive relevant doses. The level of details needed depends on the objective that is pursued.

When large data are generated, they can feed into *in silico* models [31].

12.2.3 *In Vitro* 2D Assays

Several 2D assays are currently available. They may be based on primary cells, cell lines, or stem cells. Cell lines are usually easy to handle and readily accessible. Attention should be paid to the genotype of the cell line that can interfere with the endpoint monitored. This is the case for genotoxicity. Primary cells can be expensive and difficult to obtain. In addition, ethical considerations may prevent the use of the latter. Stem cells hold the promise of offering a source of a variety of cell types with diversified genotypes and at different life stages. While this holds true for some cell types [32,33], a lot of effort is still ongoing to obtain fully mature cells such as stem cell-derived hepatocytes. Several R&D initiatives are ongoing in the stem cell area [34–36].

The cells can be engineered and sensors can be included in the tissue culture environment [37]. Different culture conditions are also available: static, dynamic, microfluidic devices, media with or without serum, specific gas supply, temperature, and humidity conditions. Depending on the purpose, it is important to explore the strengths, limitations, and physiological relevance of the cell/tissue type.

For *in vitro* testing, technologies such as omics (toxicogenomics, transcriptomics, proteomics, and metabolomics), high-content screening (HCS), and high-throughput screening (HTS) can help investigate the molecular mechanisms leading to specific endpoints [38–44]. They provide an insight into cellular pathways that are perturbed by the test agent. These pathways can be further investigated by follow-up assays to better characterize qualitatively and quantitatively the nature of the changes.

Coculturing different cells can be an option to address issues where the function of one cell type depends on the presence of another one. This is the case with keratinocyte–melanocyte cocultures [45].

Technologies such as toxicogenomics, HCS, and HTS can help investigate the molecular mechanisms leading to specific effects. In the literature, several studies aiming at using toxicogenomics for predicting genotoxicity and carcinogenicity can be found [46–49]. HCS-based *in vitro* micronucleus assays are available using automated image analysis platforms [50,51]. Combining different *in vitro* approaches such as toxicogenomics, HCS, HTS, and binding to nuclear receptors holds great promise in terms of unraveling molecular events. This is one of the goals of the US EPA ToxCast program [52]. However, analyzing the large amount of data generated in this type of project is a real challenge and collaborative effort is needed to do so.

There is a need for a toolbox of models/methods to choose from depending on the endpoint/mechanism that is addressed. A variety of biological models are

now available: primary cells, cell lines, and stem cells (embryonic and induced pluripotent stem cells (IPSCs)).

12.2.4 Organotypic Models

They range from spheroids, reconstructed tissues, micropatterned cells, bioprinted tissues to more sophisticated models such as microfluidic bioreactors. In most cases, the culture conditions of organotypic models provide them with additional features that are not present under two-dimensional conditions (long shelf life and functionality of some enzymes). These models are used to perform assays where a response closer to a physiological situation is needed. The American Interagency Partnership on Tissue on a Chip (NCATS, DARPA, and FDA) is an example of such an initiative that aims at eliminating toxic/ineffective drugs early in the drug development process [53].

Building such complex models requires the multidisciplinary expertise of bioengineers, biologists, and toxicologists [54]. Rather than collecting and integrating pieces of information from different assays, the results from organotypic models provide information taking into account the complexity of the system. In some cases, they can bring added value in a weight-of-evidence approach.

As these models can retain their phenotype and functionality over several weeks, they can be used to perform repeated dosing studies *in vitro*. They are useful for performing toxicity assays (related to cell viability or vital cellular function [55–57]) or assays on ADME issues [58–60]. The ADME data can be used to derive toxicokinetic information [61,62]. The latter is useful to extrapolate information from *in vitro* to *in vivo* [63–65]. At the moment, the vast majority of these models are in a research and development stage.

Results obtained using these models can help bridge the gap between *in vitro* and *in vivo*.

12.3 Genotoxicity as a Case Study[2)]

Genotoxicity is usually addressed early in the safety assessment process of chemicals for regulatory purpose. Gene mutation and chromosomal damage are the two endpoints that are usually investigated. Although several *in vitro* assays are currently available, challenges still remain in using the results derived from these assays to support risk assessment. The main reason for that is the low specificity (many *in vitro* positives are irrelevant to human) of the *in vitro* genotoxicity assays [66]. This issue has been discussed by several organizations (European Centre for the Validation of Alternative Methods (ECVAM), European Centre

2) The experimental work presented here was performed at the Institut Pasteur de Lille (IPL).

for Ecotoxicology and Toxicology of Chemicals (ECETOC), European Cosmetic Trade Association – Cosmetics Europe, International Life Sciences Institute's Health and Environmental Sciences Institute (ILSI/HESI), International Workshops on Genotoxicity Testing (IWGT), and International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)) and different projects aimed at bringing solutions are ongoing.

Recently, the Scientific Committee on Consumer Safety (SCCS) has issued an amendment to the 8th revision of its note of guidance for the testing of cosmetic substances and their safety evaluation. This document is focused on genotoxicity and carcinogenicity. A two-test battery is recommended as a starting point for testing: a bacterial reverse mutation test (OECD Guideline 471) as a test covering gene mutations and an *in vitro* micronucleus test (OECD Guideline 487) as a test for both structural (clastogenicity) and numerical (aneugenicity) chromosomal aberrations. As follow-up assays for genotoxic substances, the comet and the micronucleus assay on reconstructed tissues are seen as promising.

Different points that can potentially contribute to improving the performances of *in vitro* genotoxicity assays have been or are being addressed: choice of cell type [67,68], toxicity measurement [69], top concentration [70], and use of 3D reconstructed tissues [71–75].

To address the need for predictive methods in the context of the European legislation (7th amendment to the European Cosmetics directive, REACH), an *in vitro* micronucleus/comet assay using a human reconstructed skin model and target cells grown beneath the skin has been developed. This coculture system aims at improving the relevance of exposure conditions in *in vitro* genotoxicity assays for topically applied compounds. In this context, Episkin® (SkinEthic SNC, Lyon, France) is used as a metabolically active tissue and a physiologic barrier. Indeed, Episkin is an *in vitro* reconstructed human epidermis from normal adult human keratinocytes cultured on a composite collagen matrix at the air–liquid interface. This biological model is histologically similar to the *in vivo* human epidermis and is already used in safety assays (skin corrosion testing, skin irritation assay, and acute and chronic skin irritation for topical formulations) and in efficacy assays (skin permeability and metabolism, effects of UVA and UVB irradiation/UVB protection, bacterial adhesion for screening antibiotics, and genomic and transcriptomic signatures). Episkin was exposed to test items topically applied. At the harvest time, keratinocytes were mechanically and enzymatically isolated for the comet assay. The comet assay performed at pH $\geq$13 allows the detection of alkali-labile sites that can be converted to single-strand breaks (SSBs), incomplete base-excision repair sites, DNA–protein (interstrand) and DNA–DNA (intrastrand) cross-links, and apoptotic cells. Concurrently to the exposure to different concentrations of the different substances, TK6 cell cultures were cocultured for a period sufficient to undergo at least one mitosis for the micronucleus test.

The *in vitro* micronucleus assay is a genotoxicity test system for the detection of chemicals or physical mutagens that induce the formation of small membrane-bound DNA fragments, that is, micronuclei in the cytoplasm of interphase

cells. These micronuclei may originate from acentric fragments (chromosome fragments lacking a centromere) or whole chromosomes, which are unable to migrate with the rest of the chromosomes during the anaphase of cell division [76]. The purpose of the *in vitro* comet assay following the alkaline version (pH >13) developed by Singh *et al.* [77] is to identify agents that induce DNA damage such as SSBs or double-strand breaks (DSBs), alkali-labile sites, DNA–DNA/DNA–protein cross-linking, and SSBs associated with incomplete excision repair sites, in one or several target organs under these experimental conditions. The single-cell gel electrophoresis (SCGE) assay, also known as the "comet assay," is a rapid, simple, visual, and sensitive technique for measuring and analyzing DNA breakage in mammalian cells [77–79], even at low levels.

Concurrently to the comet assay, the nondenaturating fast halo assay (FHA) was performed; this assay allows discriminating apoptotic from nonapoptotic cells bearing DNA single strand breaks [80]. The reconstructed human skin model presents advantages for human risk assessment of a test item applied on the skin. Indeed, Episkin reconstitutes a model that is very close to human skin (histological structure and metabolic activity) and cells are more easily isolated from Episkin than from human skin. Otherwise, the production of Episkin is standardized and is more reproducible than human skin. Furthermore, test results on human model are more pertinent in the assessment of possible human hazard. Several papers were published on this model as a model for genotoxicity testing [71,72,81]. Regarding TK6 cell line, it is a continuous human lymphoblastoid cell line with a normal and stable p53 status ($p53^{+/+}$). This cell line is recommended in the OECD Guideline 487 for the *in vitro* micronucleus test [82].

In order to assess and validate the coculture model Episkin/TK6, 13 reference compounds were tested in both assays. These substances were chosen in accordance with the list published by Kirkland *et al.* [83]. Among these molecules, (1) *in vitro* and *in vivo* genotoxins were chosen: 7,12-dimethylbenzanthracene (DMBA), ethylnitrosourea (ENU), dimethylnitrosamine (DMN), and hydroquinone; (2) non-DNA-reactive chemicals (including nongenotoxic carcinogens) that have been reported to induce positive results *in vitro* (CA, MLA/TK), often at high concentrations or at high levels of cytotoxicity: curcumin, *o*-anthranilic acid, D,L-menthol, sulfisoxazole, benzyl alcohol, and ethionamide; and (3) non-DNA-reactive chemicals (including nongenotoxic carcinogens) that give negative results in *in vitro* mammalian cell genotoxicity tests: ampicillin trihydrate, D-mannitol, and phenanthrene.

Furthermore, the micronucleus assay was also performed on reconstructed human epidermises (RHE®; SkinEthic SNC, Lyon, France) on D,L-menthol.

12.3.1 Materials and Methods

12.3.1.1 Episkin

Episkin is a human epidermis reconstructed *in vitro* from normal adult human keratinocytes, cultured on a collagen matrix composite at the air–liquid interface

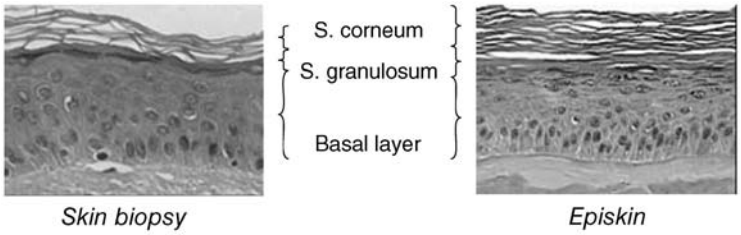


Figure 12.1 An engineered epidermis made of human keratinocytes. (Courtesy of SkinEthic Laboratories.)

(Figure 12.1). The biological model is histologically similar to human skin *in vivo* and was already used in safety testing and efficacy.

Episkin 1.2 cm^2 skin equivalents were purchased from SkinEthic SNC (Lyon, France). Upon receipt, the reconstructed skin inserts were transferred into 12-well plates containing 2 ml of Episkin "maintenance" medium and kept in an incubator at 37 °C under 5% CO_2 and 95% humidity. The following day, the medium was replaced by 2 ml of Episkin "treatment" medium according to manufacturer's instructions.

12.3.1.2 **RHE**

RHE 0.5 cm^2 skin equivalents were purchased from SkinEthic SNC (Lyon, France). Upon receipt, the reconstructed skin inserts were transferred into 24-well plates containing 1 ml of Skinethic® "maintenance" medium and kept in an incubator at 37 °C under 5% CO_2 and 95% humidity. One hour later, the medium was replaced by 1 ml of Skinethic "treatment" medium containing cytochalasin B (3 µg/ml) according to manufacturer's instructions.

12.3.1.3 **TK6 Cells**

The TK6 human B-lymphoblastoid cell line was obtained from the American Type Culture Collection (Rockville, MD). TK6 cells were isolated from a culture of the spontaneously immortalized splenic lymphoblasts WIL2 that possess a stable genome, with a normal and stable p53 status ($p53^{+/+}$).

A stock was cryopreserved in liquid nitrogen at the laboratory. Each new batch of cells was tested to confirm the absence of mycoplasma contamination.

The cells were maintained in suspension at 37 °C in a humidified 5% CO_2 atmosphere in RPMI 10 medium (RPMI 1640 growth medium (RPMI 0 medium) supplemented with 10% (v/v) heat-inactivated horse serum, 200 U/ml penicillin, 50 µg/ml streptomycin, 2.5 µg/ml amphotericin B, 200 µg/ml L-glutamine, 200 µg/ml sodium pyruvate, and 500 µg/ml pluronic acid). During the period of cell expansion, cells were maintained at an average density of 3×10^5 to 7×10^5 cells/ml by counting and diluting cells every 2–3 days. Cell culture was never left unattended for longer than 3 h and never used at a density of more than 9×10^5 cells/ml. The doubling time of TK6 cells was 16–18 h.

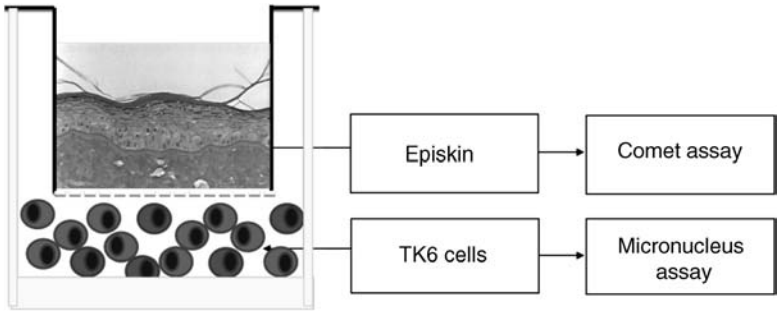


Figure 12.2 The Episkin/TK6 coculture. (Courtesy of SkinEthic Laboratories.)

12.3.1.4 Episkin + TK6 Cells Coculture

After an overnight period in maintenance medium, the Episkin inserts were placed in 12-well plates containing a fresh suspension of TK6 cells (150 000 cells/ml) in Episkin "treatment" medium (Figure 12.2).

12.3.2 Chemicals

The test chemicals are listed in Table 12.1 with the corresponding abbreviations, CAS registry numbers, sources, purities, their known genotoxic status, and IARC classification. All chemicals were dissolved in dimethyl sulfoxide (DMSO).

12.3.3 Treatment Schedules

12.3.3.1 Episkin + TK6 Cells Coculture

Two assays were implemented: in the first assay, the epidermises and the TK6 cells were harvested 4 and 54 h after the application of the test item, respectively, given that a recovery period of 27 h was observed before harvesting TK6 cells. In the second assay, the epidermises and TK6 cells were both harvested 27 h after the topical exposure. In each assay, the test item was applied at different doses with 50 µl/well on the upper side of the wells using a mesh to allow a better diffusion of the test item to the whole surface of the epidermises. After the 27 h treatment period, the TK6 cells were transferred into 15 ml tubes, centrifuged for 6 min at 1000 rpm, and poured off to discard the supernatant. The cells were washed twice with PBS. After the washing, cells were resuspended in 2 ml of RPMI 10, transferred into 12-well plates, and then incubated for 27 h (first assay S9 – 27 h/+27 h) or were harvested (second assay S9 – 27 h/+0 h).

Concurrently, wells were treated under the same conditions with the solvent only (untreated controls). Mitomycin C at a concentration of 7 µg was used as a positive control for the micronucleus assays on TK6 cells and methyl methanesulfonate (MMS) 200 and 20 µg final doses were used as positive controls for the

Table 12.1 List of chemicals tested: true positives (*in vitro* and *in vivo* genotoxins); true negatives (non-DNA-reactive chemicals (including nongenotoxic carcinogens) that give negative results in *in vitro* mammalian cell genotoxicity tests); and "irrelevant" positives (non-DNA-reactive chemicals (including nongenotoxic carcinogens) that have been reported to induce positive results *in vitro*).

Test item	CAS no.		*In vivo* genotoxic test	*In vitro* mammalian cell test	Carcinogenicity	IARC classification	Further information	References
In vivo genotoxins that should be detected as positive in *in vitro* mammalian cell genotoxicity tests								
ENU	759-73-9	+ve	+ve for CA, MN, and transgenic mutations in many tissues	+ve for MN and CA tests in the range 10–100 μg/ml	Nervous system, small intestine, and thyroid tumors in rats. Skin tumors in mice after dermal application (IARC vol. 17).	IARC group 2A carcinogen	Strong gene mutagen (O^6-alkylation)	[84–89]
7,12-Dimethylbenzanthracene	57-97-6	+ve	+ve for MN and gene mutations	+ve for MLA (+S9) at <10 μg/ml; +ve for MN; variable CA responses from 1 to 200 μg/ml (– and +S9)	Vascular tumors in female mice (not tested systemically in rats). Skin tumors in mice, hamsters, and gerbils following dermal application	Not classified by IARC with regard to human carcinogenicity	Requires metabolic activation (CYP1B1); forms bulky adducts	[66,85,87, 89–92]
Dimethylnitrosamine	62-75-9	+ve	+ve for gene mutations and UDS in liver; –ve for MN	+ve for MLA in the range 9–250 μg/ml; induces MN; also +ve for CA (+S9) in the range 500–7500 μg/ml	Liver tumors in rats and mice, but also lung, nervous system, kidney, testes, and vascular tumors	IARC group 2A carcinogen	Alkylating agent after activation by CYP2E1 (which is not highly expressed in rat liver S9): produces O^6- and N^7-methylguanine adducts	[85–87, 92–97]
Hydroquinone	123-31-9	+ve	+ve for CA (IARC vol. 71) and MN	+ve for MLA; +ve for MN and CA at <10 μg/ml	Kidney, liver, and hematopoietic tumors in rats and mice		MOA: aneugen	[85,86,93, 98,99]

(*continued*)

Table 12.1 (*Continued*)

Test item	CAS no.		*In vivo* genotoxic test	*In vitro* mammalian cell test	Carcinogenicity	IARC classification	Further information	References
Non-DNA-reactive chemicals (including nongenotoxic carcinogens) that should give negative results in *in vitro* mammalian cell genotoxicity tests								
Noncarcinogens with negative *in vivo* genotoxicity data								
Ampicillin trihydrate	7177-48-2	+ve	−ve for MN; −ve for MLA up to 5000 μg/ml; −ve for CA up to 1500 μg/ml		−ve in rats and mice			[84,85, 93,95, 100–102]
D-Mannitol	69-65-8	+ve	−ve for MN and CA	−ve for CA and MLA up to 5000 μg/ml	−ve in rats and mice			[84,85,91, 93,95,100, 103,104]
Noncarcinogens with no *in vivo* genotoxicity data								
Phenanthrene	85-01-8	−ve [86] (IARC vol. 32)	No data	−ve for CA (details not available, IARC vol. 32)	−ve in mice after dermal, i. p., and subcutaneous administration (IARC vol. 32)			[84,85,105]
Non-DNA-reactive chemicals (including nongenotoxic carcinogens), metabolic poisons, and others that should give negative results in *in vitro* mammalian cell genotoxicity tests, but have been reported to induce chromosomal aberrations or *tk* mutations in mouse lymphoma cells, often at high concentrations or at high levels of cytotoxicity								
Noncarcinogens that are negative or equivocal for genotoxicity *in vivo*								
D,L-Menthol	15356-70-4	−ve	−ve for MN	−ve/inconclusive for MLA; +ve for CA, 3 h (−S9) (+17 h recovery), 1.6–1.9 mM with toxicity	−ve in rats and mice			[93,95,100, 106–108]
o-Anthranilic acid	118-92-3	−ve	−ve for MN and CA	+ve for MLA and CA at toxic concentrations; +ve for MN *in vitro* above 4000 μg/ml	−ve in rats and mice			[93,95,100, 109–112]

+Sulfisoxazole	127-69-5	−ve	−ve for MN and CA *in vivo*	−ve for CA and but inconclusive or weakly +ve at <20% RTG MLA	−ve in rats and mice	[84,85, 95,100, 113,114]
Noncarcinogens with no *in vivo* genotoxicity data						
+Ethionamide	536-33-4	−ve	No *in vivo* genotoxicity data	Weakly +ve for MLA at 70–90% toxicity; weakly +ve for CA 5–8 mM with precipitation	−ve in rats; possible thyroid tumors in mice	[84,106]
Curcumin	458-37-7	−ve	No *in vivo* genotoxicity data for curcumin alone	+ve for MN only with apoptosis	Anticarcinogen	[84,115, 116]
+Benzyl alcohol	100-51-6	−ve	No *in vivo* genotoxicity data	Weakly +ve for MLA at 4500 µg/ml and weakly +ve for CA at 4000 µg/ml, that is, 30–40 mM	−ve in rats and mice	[84,85, 95,100]

comet assay on keratinocytes for both the short and the continuous expression time, respectively.

12.3.3.2 RHE

The treatment schedules are those described by Curren *et al.* [71].

The test item was topically applied three times at 24 h interval at different doses. Fifty microliters per well were deposited using a mesh to allow a better diffusion of the test item to the whole surface of the epidermises. The medium was replaced every day by fresh medium containing cytochalasin B.

At the harvesting time (24 h after the third treatment), cells from the basal layer and stratum spinosum of the models were harvested and prepared for slide analysis as described by Curren *et al.* [71].

For tissues, cytotoxicity was assessed by both the MTT assay and the FHA [117,118]. For the MTT technique, the inserts were transferred into 12-well plates containing 2 ml of MTT (0.5 mg/ml in medium without phenol red). The plates were then incubated for at least 90 min in a CO_2 incubator at 37 °C. After this contact time, the epidermises were punched and placed in Eppendorf tubes containing 1 ml of a mixture of isopropanol/HCl (1/23, v/v). The Eppendorf tubes were then placed at room temperature, protected from light, for one night. At the end of this period, the tubes were agitated and 2×0.1 ml per well were transferred into a flat-bottom plate. Absorbances were determined using Diagnostic Pasteur LP400 equipment at 550 nm with a reference wavelength of 620 nm. The relative toxicity was calculated for each concentration using the ratio: [(mean absorbance at the tested concentration)/(mean absorbance of the negative control)] $\times$ 100. A test item was considered cytotoxic below 50% of relative survival rate. FHA is a sensitive and reliable method to detect DNA strand breakage induced either by various genotoxic agents or secondary to apoptotic DNA cleavage [119]. In this procedure, the electrophoresis step was omitted from the comet assay procedure. After incubation in fresh electrophoresis buffer for 20 min at room temperature, the slides were directly washed in neutralization buffer without electrophoresis. This method avoids excessive migration of highly unwound DNA. DNA from cells having heavily damaged DNA appears as a "halo" and the distinction between normal, necrotic, and apoptotic cells was possible. One hundred randomly selected individual cells were analyzed on two slides per concentration (200 cells per concentration).

Cytotoxicity on TK6 cells was assessed through the determination of the relative population doubling (RPD), calculated as described in the OECD Guideline 487: [(number of population doublings (PD) in treated cultures)/(number of PD in control cultures)] $\times$ 100, with PD = [log(concentration of harvested cells/concentration of treated cells)]/log 2. The maximum concentration generally produces a RPD $\geq$50 $\pm$ 5%. In RHE cells, the percentage of binucleated cells (OECD Guideline 487) at each dose of test chemical is compared with the solvent or untreated control. The highest concentration to be analyzed should not produce no more than a 70% decrease in binucleated cells compared with the control.

12.3.3.3 Micronucleus Assay

In Vitro Micronucleus Assay on TK6 Cells

At the end of the treatment period, the cells were transferred into 15 ml tubes and were centrifuged for 6 min at 1000 rpm. The supernatants were hence removed, the cell pellets loosened with agitation, and 2 ml of a mixture of RPMI 0/distilled water (1/1, v/v) was slowly added. After 4 min at ambient temperature, 0.5 ml of cold ethanol/acetic acid mixture (3/1, v/v) was slowly added to pre-fix the cells. The cell suspensions were centrifuged at 1000 rpm for 6 min, the supernatants were removed, and 2 ml of cold ethanol/acetic acid mixture was added under agitation. The cell suspensions were stored at 2–8 °C for one night and then centrifuged at 1000 rpm for 6 min. Each supernatant was removed, the cell pellets loosened, and a sufficient volume of cold ethanol/acetic acid mixture was added. To end, cell suspension was dropped onto a clean, dry microscope slide. Two slides were prepared from each culture. Once completely dried, slides were stained with Giemsa (4% in mineralized water) for 10 min. After coding slides, the cells were examined under the microscope and screened for micronuclei in mononucleated cells (MNCs). The micronuclei of at least 1000 mononucleated cells per slide were counted (2000 mononucleated cells/culture).

In Vitro Micronucleus Assay on RHE

Tissue-containing inserts were removed from the treatment plates, their bottoms blotted to remove excess medium, and placed into individual wells (each containing 5 ml PBS) of a 12-well plate for 5–15 min at room temperature.

The inserts were then transferred (after blotting) to another well (containing 5 ml EDTA 0.1%, at room temperature) in a 12-well plate and allowed to sit for 15 min.

The inserts were once more blotted and transferred to another well of six-well plates, which contain 1 ml warm (37 °C) trypsin–EDTA solution. An additional 0.5 ml of the same solution was added to the top surface, and the tissues were allowed to sit for 15 min at room temperature.

Each tissue was carefully separated from the supporting membrane by lifting the edge of the tissue with fine forceps, transferred to a new well in six-well plates, and exposed to 1 ml of fresh trypsin–EDTA for 5 min. To capture cells still adherent to the membrane of the insert, the insert (containing any remaining trypsin–EDTA) was placed in the same well as its matching tissue.

After 5 min, the insert was thoroughly rinsed (four to six times) to collect the trypsinized cells. The tissue was then agitated to release any remaining attached cells, and the resulting cell clumps and tissue additionally disrupted by repeatedly drawing into a pipette and expelling the solution. The single-cell suspension (1.5 ml) was transferred into a 15 ml conical tube containing 8.5 ml of warm MEM with 10% FBS to neutralize the trypsin.

A 1 ml sample of cells was diluted with 1 ml trypan blue solution (or less, depending on sample volume) and counted using a hemocytometer. The

remaining cell suspension was centrifuged at $100 \times g$ for 5 min. After the centrifugation, the supernatant was carefully removed, the cell pellet loosened with agitation, and 1 ml of warm (37 °C) KCl (0.075 M) solution was slowly added. 0.5 ml of cold methanol/acetic acid (3 : 1) was added immediately. After 1 min, 3 ml of cold methanol/acetic acid (3 : 1) was added to fix the cells, and the solution mixed. The cell suspension was centrifuged at $100 \times g$ for 5 min. The supernatant was removed, the pellet loosened, and 3 ml of cold methanol/acetic acid (3 : 1) was added. All but a small portion of the supernatant was then removed, the cell pellet loosened, and a drop of the cell suspension was pipetted onto a clean, dry microscope slide. One slide was prepared from each tissue. Once completely dry, slides were immersed in AO solution (100 µg/ml) for 10 min. Slides were then scored using a fluorescence microscope. The micronuclei of at least 500 binucleated cells per culture were counted (1000 binucleated cells/dose). Micronuclei were identified according to the criteria of Fenech [119,120].

Statistical analysis of the results obtained in the cells treated at each dose level was performed using the χ^2-test in comparison with those in control groups.

12.3.3.4 *In Vitro* Comet Assay Protocol

At the end of the treatment, transwells were removed and the epidermises were punched. The epidermis was then mechanically detached from the collagen support and placed in a new 12-well plate. The epidermises were thoroughly scraped in 1 ml of trypsin/EDTA 0.125% (previously heated at 37 °C) and placed in a cell incubator for 20 min at 37 °C under 5% CO_2 with 95% humidity. Agitation was regularly done. At the end of this period, 0.5 ml of trypsin/EDTA 0.125% was added, and the plates were placed in a cell incubator for 10 min at 37 °C. The enzymatic digestion was stopped by adding 300 µl of fetal calf serum (Sigma). The resulting cell suspension was transferred into a 15 ml tube and centrifuged for 6 min at 1000 rpm. The cell pellet was then ready for the cytotoxicity assay and/or for the comet assay procedure.

The comet assay was performed as previously described by Singh *et al.* [77] with slight modifications. Briefly, in a light-free room, 5×10^4 cells were suspended in 150 µl of molten 0.5% low melting point agarose (LMPA; Bio-Rad, France) in PBS without calcium or magnesium. Seventy-five microliter aliquots of the cell suspension were rapidly spread on each of the two slides (Touzard et Matignon, Courtaboeuf, France) precoated with 85 µl normal agarose (0.8% in Ca^{2+}/Mg^{2+}-free PBS) and then covered with coverslips (24 mm × 32 mm). After the agarose was allowed to solidify for 10 min at 0 °C, the coverslips were removed. The slides were then placed in a tank filled with lysis solution (2.5 M NaCl, 0.1 M EDTA, pH 10, 10% DMSO, and 1% Triton X-100 both freshly added) at +4 °C. After 1 h, the slides were removed from the lysis solution and incubated in fresh electrophoresis buffer (0.3 M NaOH and 1 M EDTA, pH >13) for 20 min at room temperature for unwinding of DNA. Electrophoresis was then carried out at room temperature in the same electrophoresis buffer for 20 min at 0.7 V/cm and 300 mA. After electrophoresis, the slides were washed

twice for 10 min in fresh neutralization buffer (0.4 M Tris–HCl, pH 7.5), followed by dehydration in absolute ethanol.

The slides were next stained with 20 µl of propidium iodide (20 µg/ml) and covered with a coverslip. For each treatment concentration, 50 randomly chosen cells from each of two slides (100 cells per culture, 200 cells per dose) were analyzed. Slides were examined with a 200× magnification, using a fluorescent microscope (Leica Microsystems SAS – DM 2000, Heerbrugg, Switzerland), equipped with an excitation filter of 515–560 nm and a barrier filter of 590 nm, connected through a gated monochrome CCD IEEE1394 FireWire video camera (Allied Vision Technologies) to the Comet Assay IV Image Analysis System, version 4.11, with Windows XP Pro Software (Perceptive Instruments Ltd, Suffolk, UK).

Ghost cells morphology was indicative of highly damaged cells often associated with severe genotoxicity, necrosis, and apoptosis. They were identified visually, as nuclei with small heads and large tails located at a distance from the head, as described by Fairbairn *et al.* [121], Kizilian *et al.* [122], and Olive *et al.* [78], and were enumerated independently.

12.3.3.5 Statistical Analysis

Before analyzing the median of the Olive tail moment (OTM) of the treated and control groups, the absence of statistically significant differences between group variances was verified. The group variances were compared using the F of Snedecor at the 0.05 significance level. When differences between group variances were not found to be significant, the parametric, one-way ANOVA test was performed on the median of OTM. The parametric t-test was then applied to compare each of the doses tested with the vehicle control in order to determine statistical significance of differences in group median values between each group and the solvent control. When differences between group variances were found to be significant, the one-way Kruskal–Wallis nonparametric test was performed on the median of OTM. The nonparametric Mann–Whitney U-test was then applied to compare each of the doses tested with the solvent control.

Statistical analysis of the results obtained for the number of ghost cells was performed using the χ^2-test in comparison with the control group.

12.3.4 Results

12.3.4.1 *In Vitro* and *In Vivo* Genotoxins

Table 12.2 summarizes the results obtained in the Episkin/TK6 coculture system by performing the micronucleus test on TK6 cells and the comet assay in keratinocytes isolated from the tissues.

ENU or DMBA induced statistically significant levels of DNA primary damage in keratinocytes harvested 4 or 27 h after their application, as

Table 12.2 Summary of the results obtained in the Episkin/TK6 coculture system by the means of the micronucleus test in TK6 cells and the comet assay in keratinocytes dissociated from the tissues.

Compound	Dose (μg/epidermis)	Assay 1					Assay 2				
		In vitro micronucleus assay on TK6 cells (27 h/+27 h)		Comet assay on Episkin (4 h/+0 h)			*In vitro* micronucleus assay on TK6 cells (27 h/+0 h)		Comet assay on Episkin (27 h/+0 h)		
		% RPD[a]	No. MMNCs/ 2000 MNCs	% Necrosis[b]	% Ghost cells	Median OTM[c]	% RPD	No. MMNCs/ 2000 MNCs	% Necrosis	% Ghost cells	Median OTM
In vitro and *in vivo* genotoxins											
ENU	0	—	6	16	37.8	2	—	6	28.5	24.6	0.44
	6.25	86.1	10	—	—	—	—	—	—	—	—
	12.5	76.4	16*	—	—	—	90	9	19.5	11.1***	2.3*
	25	68.4	10	—	—	—	84	14	24	16.4*	3.6*
	50	—	—	—	—	—	90.6	17*	32.5	34.1*	13.2**
	62.5	—	—	24	37.8	11.2**	—	—	—	—	—
	125	—	—	21	35.2	24.5**	—	—	—	—	—
	250	—	—	23	41	25.2**	—	—	—	—	—
	Mito C[d]	45.5	129***	—	—	—	80.7	47***	—	—	—
	MMS[e]	—	—	35.5***	41.3	16.77**	—	—	41**	37.1***	22.53**
DMBA	0	—	6	9	0.6	0.67	—	6	28.5	24.6	0.44
	312.5	77.7	10	—	—	—	90	12	24	25.8	3*
	625	78.2	16*	12.5	16.5***	1.3	92.6	17*	17**	23.8	10.4**
	1250	70.2	9	9.5	31.5***	2.4**	85.7	11	17.5**	24.5	11.2**
	2500	—	—	11.5	22.6***	2.8**	—	—	—	—	—
	4000	—	—	8.5	—	1.5**	—	—	—	—	—
	Mito C[d]	45.5	129***	—	—	—	80.7	47***	—	—	—
	MMS[e]	—	—	12.5	30.4***	28.5**	—	—	41**	37.1	22.53**

DMN	0	—	6	11	36.8	2	—	8	7.5	25.8	0.01
	1250	127	10	22**	25.5	1.8	111.4	18*	10	32.9	0.33*
	2500	132.4	11	16.5	34.6	1.9	103.6	5	10.5	31.6	0.07
	5000	133.2	7	15	33.5	2.3	89.8	15	9.5	20.1	0.04
	Mito C[d]	-9.6	199***	—	—	—	43.7	54***	—	—	—
	MMS[e]	—	—	18*	40.7	16.77***	—	—	19.5**	28.3	14.1*
Hydroquinone	0	—	6	13.5	31.3	2.8	—	8	8.5	25.8	0.5
	2.5	—	—	—	—	—	163.8	10	—	—	—
	5	97.1	6	—	—	—	92.7	6	—	—	—
	10	96.6	10	—	—	—	56.7	7	—	—	—
	20	48.5	21**	—	—	—	—	—	7	14.5**	1
	40	—	—	14	39.2***	18.9**	—	—	8	21.4	1
	80	—	—	13.5	39.7***	19.2**	—	—	9.5	23.1	1
	160	—	—	16	42.4***	20.2**	—	—	6.5	30	1.4**
	Mito C[d]	22.5	84***	—	—	—	43.7	54***	—	—	—
	MMS[e]	—	—	11	28.2*	21.46***	—	—	22***	28.3	24.5***
Non-DNA-reactive chemicals (including nongenotoxic carcinogens) that give negative results in *in vitro* mammalian cell genotoxicity tests											
Noncarcinogens with negative *in vivo* genotoxicity data											
Ampicillin trihydrate	0	—	9	18.5	29.8	0.8	—	13.5	10.5	20.5	0.2
	1250	86.1	2	19	28.6	0.6	90	12	12.5	24.4	0.5
	2500	84.6	8	10*	28.6	0.6	84	5*	9	22.2	0.3
	5000	78.2	6	8.5**	34.1	0.7	90.6	9	9.5	19.6**	0.3
	Mito C[d]	45.5	171***	—	—	—	80.7	46***	—	—	—
	MMS[e]	—	—	35.5***	24.5	6.82***	—	—	11	30.9**	5.83*
D-Mannitol	0	—	6	13.5	21.3	2.8	—	13.5	10.5	20.5	0.4
	1250	101.1	9	8	35.1***	5.9**	113.4	11	8	30.5**	0.3
	2500	108.6	4	11.5	33.8***	12.2*	107.9	15	7	28.4**	0.2
	5000	97.5	1	12	23.6	2.5	122.7	8	6.5	29.4**	0.3
	Mito C[d]	22.5	84***	—	—	—	82.5	46***	—	—	—
	MMS[e]	—	—	11	28.2*	21.46***	—	—	11	30.9**	22.53***

(*continued*)

Table 12.2 (*Continued*)

Compound	Dose (µg/epidermis)	Assay 1					Assay 2				
		In vitro micronucleus assay on TK6 cells (27 h/+27 h)		Comet assay on Episkin (4 h/+0 h)			*In vitro* micronucleus assay on TK6 cells (27 h/+0 h)		Comet assay on Episkin (27 h/+0 h)		
		% RPD[a)]	No. MMNCs/ 2000 MNCs	% Necrosis[b)]	% Ghost cells	Median OTM[c)]	% RPD	No. MMNCs/ 2000 MNCs	% Necrosis	% Ghost cells	Median OTM
Noncarcinogens with no *in vivo* genotoxicity data											
Phenanthrene	0	—	5.5	11	36.8	2	—	13.5	10.5	20.5	0.22
	1250	104.8	10	7.5	35.5	1*	127.6	11	7.5	22.3	0.06*
	2500	102.6	10	6.5	34.6	1*	138	6	5.5	24.8	0.18
	5000	94.1	8	7	33.5	0.7*	102.2	11	10	28.3*	0.05*
	Mito C[d)]	9.6	199***	—	—	—	82.5	46***	—	—	—
	MMS[e)]	—	—	18*	40.7	15.96*	—	—	11	30.9**	5.83**
Non-DNA-reactive chemicals (including nongenotoxic carcinogens) that have been reported to induce positive results *in vitro* (CA, MLA/TK), often at high concentrations or at high levels of cytotoxicity											
Noncarcinogens that are negative or equivocal for genotoxicity *in vivo*											
D,L-Menthol	0	—	8	11	36.8	1.96	—	6	14	24.5	0.44
	12.5	—	—	4.5*	43.6*	1.05*	—	—	—	—	—
	25	—	—	6.5	39.4	1.13*	—	—	—	—	—
	50	—	—	8	49.7***	1.27	—	—	—	—	—
	60	—	—	—	—	—	—	—	9	15.1**	0.53
	122	91.8	9	—	—	—	85.7	13	12	18.9	0.6
	195.3	84.2	8	—	—	—	94.1	17*	12.5	—	—
	312.5	81.9	6	—	—	—	76.1	6	14.5	—	—
	Mito C[d)]	62	171***	—	—	—	80.7	47***	—	—	—
	MMS[e)]	—	—	18*	40.7	15.96**	—	—	11	37.1***	22.53***

o-Anthranilic acid	0	—	3	6	35.6	1.4	—	6	26	43.6	1.4
	312.5	92.3	4	—	—	—	84.5	5	—	—	—
	625	88.1	4	9	26.1*	2.1*	99.6	10	18.5	32.6**	4.7*
	1000	72.9	10	—	—	—	—	—	—	—	—
	1250	—	—	8.5	24.5**	3.6*	78.5	13	26	31.1**	5.1*
	2500	—	—	9.5	42	39.6*	—	—	21	35.3*	3.9*
	Mito C[d]	62.1	63***	—	—	—	75.2	65***	—	—	—
	MMS[e]	—	—	8.5	41.1	33.36*	—	—	24	28***	11.88*
Sulfisoxazole	0	—	—	—	—	—	—	10	7	35.4	0.7
	1250	—	—	—	—	—	65	17	11	42.2	2.1**
	2500	—	—	—	—	—	69.8	15	9	34.1	1.7*
	5000	—	—	—	—	—	77.7	18	11.5	41	1.4
	Mito C[d]	—	—	—	—	—	56	48***	—	—	—
	MMS[e]	—	—	—	—	—	—	—	11.5	37.1	12.94*
Noncarcinogens with no *in vivo* genotoxicity data											
Ethionamide	0	—	—	—	—	—	—	10	7	35.4	0.7
	312.5	—	—	—	—	—	81.6	17	—	—	—
	625	—	—	—	—	—	57.5	11	—	—	—
	1250	—	—	—	—	—	56.3	17	5	25.2	2.9*
	2500	—	—	—	—	—	—	—	6	29.7	2.1*
	5000	—	—	—	—	—	—	—	4.5	35.3	2.9*
	Mito C[d]	—	—	—	—	—	56	48***	—	—	—
	MMS[e]	—	—	—	—	—	—	—	11.5	37.1	12.94*
Curcumin	0	—	6	10	34.2	0.9	—	6	26	43.6	1.4
	156.25	—	—	6	24**	8.1*	—	—	15	38.7	2.1
	312.5	—	—	8	20.3***	6.6*	—	—	14	36.4	1.5
	625	—	—	8.5	17.3***	3.2	—	—	13.5	36.3	1.9
	1250	96.3	12	—	—	—	76.3	8	—	—	—
	2500	90.9	11	—	—	—	93	11	—	—	—

(*continued*)

Table 12.2 (*Continued*)

Compound	Dose (μg/epidermis)	Assay 1					Assay 2				
		In vitro micronucleus assay on TK6 cells (27 h/+27 h)		Comet assay on Episkin (4 h/+0 h)			*In vitro* micronucleus assay on TK6 cells (27 h/+0 h)		Comet assay on Episkin (27 h/+0 h)		
		% RPD[a]	No. MMNCs/ 2000 MNCs	% Necrosis[b]	% Ghost cells	Median OTM[c]	% RPD	No. MMNCs/ 2000 MNCs	% Necrosis	% Ghost cells	Median OTM
	5000	95.1	14	—	—	—	72.1	8	—	—	—
	Mito C[d]	37.5	129***	—	—	—	75.2	65***	—	—	—
	MMS[e]	—	—	7.5	27.6	18.42*	—	—	24	28***	11.88*
Benzyl alcohol	0	—	—	—	—	—	—	10	7	35.4	0.7
	625	—	—	—	—	—	86	25*	13*	32	6.5*
	1250	—	—	—	—	—	79	21*	14.5*	27.8*	4.3*
	2500	—	—	—	—	—	66.9	20	8.5	44.9*	13.9*
	Mito C[d]	—	—	—	—	—	56	48***	—	—	—
	MMS[e]	—	—	—	—	—	—	—	11.5	37.1	12.94*

a) Calculated through the trypan blue exclusion dye technique.
b) Determined through the fast halo assay. In any case, no statistically significant increase in the percentage of cells in apoptosis was noted.
c) Median calculated on 200 cells.
d) Mitomycin C: 7 μg/epidermis.
e) 20and 200 μg/epidermis for the continuous or the short expression time, respectively.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

revealed by the alkaline comet assay, with values for the medians of OTM reaching 25.2 with 250 μg of ENU/epidermis in the 4 h expression time versus 2 in the control and 13.2 at the dose of 50 μg of ENU/epidermis in the long expression time versus 0.44 in the control. The application of DMBA gave values for the medians of OTM of 2.8 and 11.2 versus 0.67 and 0.44, respectively, in the short and long expression times at 2500 and 1250 μg of DMBA/epidermis, respectively.

Furthermore, statistically significant increases in the number of micromononucleated cells (MMNCs)/1000 MNCs were also observed: 16 and 17 MMNCs/2000 MNCs were observed in TK6 cells 27 h after the application of either 12.5 or 50 μg of ENU/epidermis, respectively, in the first and the second assay. The dose of 625 μg of DMBA/epidermis induced the appearance of 16 and 17 MMNCs/2000 MNCs in TK6 cells 27 h after the application followed or not by a recovery period, respectively, versus 6 in the respective controls.

DMN assessed in the comet assay induced a slight increase in the median of OTM after a long expression time, only at the lowest dose assessed of 1250 μg/epidermis with a value for the median of OTM of 0.33 versus 0.01 in the relative control. This effect was considered as equivocal and should be confirmed at lower doses. DMN also induced a statistically significant increase in the number of micronucleated TK6 cells when using a long-term exposure, without recovery period with 18 MMNCs/2000 MNCs at 1250 μg/epidermis versus 8 in the respective control.

The comet assay performed 4 h after exposure to hydroquinone appeared to be more efficient in the *in vitro* genotoxic assay with a maximum value for the median OTM of 20.2 at 160 μg/epidermis versus 2.8 in the untreated control. When considering the results of the micronucleus test on TK6 cells, 27 h after topical application of hydroquinone followed by a 27 h recovery period, the genotoxic potential of the this substance was highlighted with a maximal value of 21 MMNCs/1000 MNCs at 20 μg/epidermis. In contrast, when the micronucleus assay was performed 27 h after application without recovery period, no significant effect was seen.

12.3.4.2 Non-DNA-Reactive Chemicals (Including Nongenotoxic Carcinogens) That Give Negative Results in the *In Vitro* Mammalian Cell Genotoxicity Tests

As expected, with ampicillin and phenanthrene, in the alkaline comet assay, no primary DNA damage was observed 4 or 27 h after their application. Moreover, no statistically significant increase in the frequency of micronucleated TK6 cells was seen in any of the assays. A statistically significant decrease, with no meaning in terms of genotoxicity, was noted 27 h after the application of ampicillin (2500 μg/epidermis).

In the comet assay performed with D-mannitol, a slight but statistically significant increase in the DNA fragmentation was noted 4 h after the topical application. However, when increasing the time of expression, no DNA damage was observable. Therefore, D-mannitol was considered nongenotoxic. In the *in vitro* micronucleus test, D-mannitol was not genotoxic either.

12.3.4.3 Non-DNA-Reactive Chemicals (Including Nongenotoxic Carcinogens) That Have Been Reported to Induce Positive Results *In Vitro* (CA, MLA/TK), Often at High Concentrations or at High Levels of Cytotoxicity

o-Anthranilic acid demonstrated genotoxicity in the *in vitro* comet assay on keratinocytes, either 4 or 27 h after application at the doses analyzed from 1000 to 2500 μg/epidermis. However, the biological relevance of the results remains to be demonstrated, as it is noteworthy that the doses retained for comet analyses were found to be toxic for the TK6 cells. Furthermore, when increasing the duration of expression in the comet assay up to 27 h, the intensity of genotoxic activity decreased. In particular, at the highest dose assessed (2500 μg/epidermis), the values of the medians of OTM decreased from 39.6 to 3.9 in the 4 and 27 h expression times, respectively.

In the *in vitro* micronucleus test on TK6 cells, no clear positive results with *o*-anthranilic acid were seen either with or without recovery period at doses ranging from 312.5 to 1000 μg/epidermis.

Curcumin was considered as genotoxic in the *in vitro* comet assay on epidermises, exclusively following a short expression time. When increasing the duration of expression, no DNA damage was observed anymore. Indeed, whereas the value for the median at the dose of 156.25 μg/epidermis of OTM reached a maximal value of 8.1 in the short expression time, it decreased to 2.1 in the long expression time, versus 0.9 and 1.4 in the corresponding untreated controls. Furthermore, an equivocal to slight genotoxic activity was noted in the *in vitro* micronucleus test on TK6 cells, only when applying a recovery period; a maximum of 14 MMNCs/2000 MNCs were observed versus 6 in the corresponding solvent control.

Sulfisoxazole, ethionamide, and benzyl alcohol were investigated in the second step using the most relevant schedule, that is, harvesting the cells 27 h after topical exposure for both the micronucleus test and the comet assay. The three substances were genotoxic in the comet assay with values for the medians of OTM reaching top values of 2.1, 2.9, and 13.9, respectively, versus 0.7 in the negative control. In contrast, in the micronucleus test on TK6 cells, no genotoxic activity was noted in presence of sulfisoxazole or ethionamide (maximum of 18 and 17 MMNCs/2000 MNCs versus 10 MMNCs/2000 MNCs in the negative control). Benzyl alcohol was considered as genotoxic in the micronucleus assay, with 20–25 MMNCs/2000 MNCs.

D,L-Menthol induced a statistically significant increase in the number of MMNCs with a total of 17 MMNCs, only at the intermediary dose assessed of 195.3 μg/epidermis, versus 6 in the respective solvent control in the *in vitro* micronucleus test following a 27 h treatment without recovery period. This effect was not observed in the second assay, without recovery period. No genotoxic activity was demonstrated in the comet assay on keratinocytes whatever the expression time [93]. As no increase in the DNA damage was observed with D,L-menthol while the micronucleus assay on TK6 cells was equivocal, the micronucleus assay was implemented on epidermal cells isolated from the tissues. In this assay, three applications at 24 h intervals were done, following the

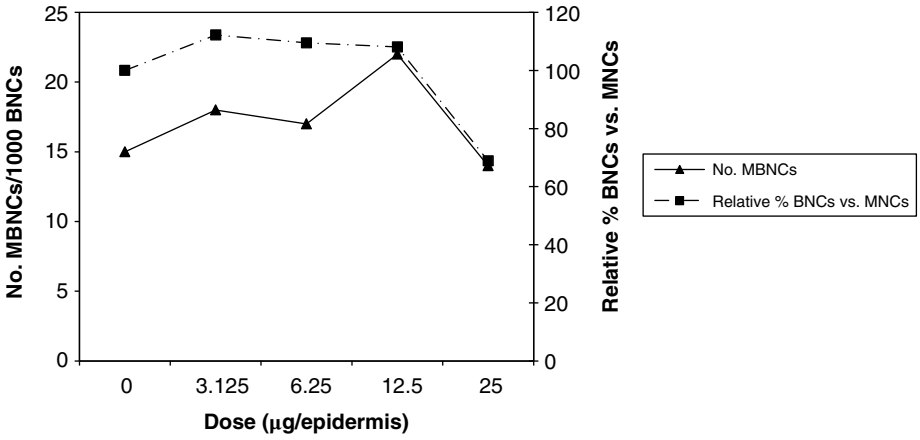


Figure 12.3 D,L-Menthol: micronucleus assay on epidermal cells isolated from the tissues.

protocol by Curren *et al.* [71]. The results are presented in Figure 12.3. No statistically significant increase in the number of micronucleated keratinocytes among binucleated cells was noted at the doses assessed from 3.125 to 25 μg/tissue.

12.3.4.4 Discussion

The current validation study was performed with the use of two complementary endpoints, that is, the chromosomal aberrations by the means of the micronucleus test in human TK6 cell line and primary DNA damage with the implementation of the alkaline comet assay in isolated keratinocytes using a coculture system. The first step was carried out on 10 compounds with different classes of genotoxicity: (1) *in vitro* and *in vivo* genotoxins, (2) non-DNA-reactive chemicals (including nongenotoxic carcinogens) that give negative results in *in vitro* mammalian cell genotoxicity tests, or (3) non-DNA-reactive chemicals (including nongenotoxic carcinogens) that have been reported to induce positive results *in vitro* (CA, MLA/TK), often at high concentrations or at high levels of cytotoxicity.

Regarding the first class of compounds, ENU, DMBA, DMN, and hydroquinone, the Episkin/TK6 coculture system was able to reveal the genotoxic activity of all these well-known *in vitro* and *in vivo* genotoxic substances. Globally, the sensitivity of the current model was better evidenced when using a long-term expression time without recovery period, as particularly demonstrated by the results obtained with DMN. In return, the micronucleus test revealed the genotoxic activity of hydroquinone when applying a longer recovery period. Specificity was also demonstrated toward nonreactive DNA compounds ampicillin, phenanthrene, and D-mannitol. As expected, no positive results were obtained, except for mannitol that induced a slight increase in DNA fragmentation after the 4 h expression time. However, this effect was not considered as relevant as it was not reproduced after the 27 h recovery period. An increased

expression time in the comet assay on keratinocytes gives more reliable results as noted with D-mannitol.

When analyzing the data gathered with D,L-menthol, *o*-anthranilic acid, or curcumin, the specificity of the coculture system was also evidenced when using a 27 h expression time, for both endpoints. When focusing on each experimental condition, it is noteworthy that *o*-anthranilic acid turned out to be genotoxic in the *in vitro* comet assay on keratinocytes, following either a 4 or a 27 h expression period. As for curcumin, it was considered as genotoxic in the *in vitro* comet assay on epidermises, exclusively following a short-term expression period. However, it is well accepted that *o*-anthranilic acid is not carcinogen *in vivo* [100], and it was previously demonstrated that the *in vitro* genotoxic activity of *o*-anthranilic acid is rather attributed to its toxicity than to an intrinsic genotoxic potential. Indeed, only high concentrations and/or toxic concentrations induce positive response in regulatory assays (i.e., *in vitro* micronucleus test, MLA/TK, or chromosomal aberrations) [93,109,110]. On the other hand, it is well known that curcumin induces positive response through apoptosis [115]: this effect is considered as an irrelevant positive event in genotoxicity tests. Therefore, when studying a 27 h expression time in both the *in vitro* comet and micronucleus assays, the Episkin/TK6 coculture system provides reliable outcomes about the genotoxic potential of curcumin and *o*-anthranilic acid.

The study using the micronucleus test on TK6 cells performed 27 h after application of D,L-menthol revealed a weak genotoxic activity. However, D,L-menthol is not carcinogenic *in vivo* [100], and several papers demonstrated that the *in vitro* genotoxicity observed in the presence of D,L-menthol is rather attributed to its toxicity than to an intrinsic genotoxic potential [106]. Indeed, only high concentrations and/or toxic concentrations induced positive or inconclusive responses in regulatory assays (i.e., MLA/TK [93] or in the chromosomal aberration assay [106]). Therefore, it seems that the comet assay on keratinocytes following a 27 h expression time demonstrated a better specificity toward this compound, rather than the micronucleus test. Under these conditions, and always in the perspective of developing alternative models for cosmetic purposes, D,L-menthol gave unexpectedly positive results in the micronucleus test on TK6 cells after a 27 h treatment. It was assessed for its genotoxic potential in another skin model, that is, RHE. This model presents the advantage of having dividing cells allowing the implementation of the micronucleus test on keratinocytes isolated from the tissues. The assessment of the genotoxic potential of D,L-menthol by the micronucleus test in this model yielded clear negative data. This promising result warrants further investigation.

When considering the overall results, one out of the four *in vivo* genotoxins (i.e., DMN) that should be detected as positive in the *in vitro* mammalian cell genotoxicity tests and one out of the three non-DNA-reactive chemicals (i.e., D-mannitol) that should give negative results in the *in vitro* mammalian cell genotoxicity tests need a long-term expression time to be evidenced. The 27 h expression time demonstrated a better specificity with regard to all the three

non-DNA-reactive chemicals (i.e., D,L-menthol, *o*-anthranilic acid, and curcumin). In the second step of the current validation study, the 27 h expression time, which seems to be the most promising and specific schedule of treatment, was used to assess the genotoxic potential of sulfisoxazole, ethionamide, and benzyl alcohol. These three compounds that are non-DNA-reactive chemicals should give negative results *in vitro* [100]. The complementary data in the micronucleus test on TK6 cells following a 27 h treatment demonstrated once again the relevance of the coculture system Episkin/TK6 under these experimental conditions. Otherwise, the results on sulfisoxazole and ethionamide met the previous data that demonstrated only weakly positive or inconclusive response in the MLA/TK assay and negative response in the chromosomal aberration test (sulfisoxazole), or positive response in the chromosomal aberration test, but at concentrations that induced toxicity or precipitation (ethionamide). In contrast, benzyl alcohol remained clearly genotoxic in both endpoints, whereas in further studies [84] authors concluded it as an irrelevant positive compound with a weak mutagenic activity in MLA/TK and chromosomal aberration tests, only at high concentrations. However, when tested on another target organ using the coculture model, both a local (on epidermises) and a systemic (on TK6 cells) genotoxic effect were found. Further studies should be implemented to ensure that benzyl alcohol displayed no genotoxic hazard *in vivo* when exposed via the skin. (This is clearly not an option for the cosmetic industry).

12.3.4.5 **Conclusions**

In order to improve the relevance of the *in vitro* genotoxicity assays, the use of three-dimensional reconstructed skin models appeared as a good complement. It may be considered as a follow-up to the current existing models for *in vitro* genotoxicity assays that present gaps regarding the relevance to human extrapolation, with numerous irrelevant positive results, whereas *in vivo* genotoxicity assays do not demonstrate any alert. Indeed, reconstructed tissues are physiologically and structurally similar to human skin and metabolically active. Otherwise, they are already successfully used in models for skin corrosivity and skin irritation, and ought to be used for genotoxicity testing as relevant results were noted in the current validation study. Therefore, the assessment of the genotoxic potential of a product could be done through the comet assay on Episkin and the micronucleus test on TK6 cells in the coculture system upon the optimal treatment schedule (i.e., 27 h treatment for both endpoints), and if necessary (e.g., positive results in only one out of the two endpoints), in a second intent, the implementation of the micronucleus test on RHE. However, data on the concentrations of tested products in culture medium, that is, in contact with TK6 cells, should be determined to know the actual level of test item and to evaluate the barrier capacity of the epidermises and its metabolic capacity. Furthermore, one should pay attention to the relevance of the results at toxic doses. In particular, data from the comet assay should be carefully analyzed especially when toxicity is demonstrated on TK6 cells at the same dose levels or if clear cytotoxicity occurs at expression time.

12.3.4.6 Related Initiatives

At Cosmetics Europe, different research projects covering the genotoxic endpoints are being performed since 2007. One of them addressed the improvement of the existing assay by investigating the cell type, the cytotoxicity assays, and human relevance of cell type [67–69]. Another one is based on the use on 3D skin models to perform the comet assay and the micronucleus assay [75,123]. As metabolism can play an important role in genotoxicity, the metabolic capabilities of the tissues were also investigated [124,125]. As stated previously, Cosmetics Europe is committed not only to developing predictive methods, but also to fostering their acceptance and use [126].

12.4 The Way Forward: Combining *In Silico* and *In Vitro* Tools

Another way to tackle this issue would be to combine *in silico* and *in vitro* tools. It is worthwhile pointing to the fact that while performances of *in vitro* assays are usually checked, those of *in vivo* ones are not readily available. Similarly to *in vitro* assays, the bioassays are providing information on hazard and not risk. So, why are *in vitro* results not used for risk assessment? Exposure data, toxicokinetics, and proper metabolic capability are lacking in simple cell systems. These are some of the types of information needed in addition to *in vitro* genotoxicity assays in order to perform risk assessment. The ILSI's Genetic Toxicity Technical Committee (GTTC) is currently investigating ways to obtain quantitative information for genotoxicity [127].

Most of the models (including the *in vivo* ones) do not provide information about the sequence of events leading to the observed endpoints: they are descriptive. What has been identified as key to using alternative methods is getting insight into mechanisms [128]. According to this report, risk assessment can become time and cost effective, if we move from descriptive to predictive (mechanistically based) toxicology. One such application has been recently described by Adeleye *et al.* [129]. The approach is AOP based and exemplified with two case studies.

Huge amounts of data have been generated by several companies and agencies worldwide. Information systems have evolved and now are able to handle and process billions of data at once. The next move (which has already started) would be to share data in order to make the best use of what have been done in past and build the toxicology of the future. Data can be shared within consortia with a well-defined legal framework. Building (Q)SARs is a way to secure the knowledge generated for decades with different *in vivo* and *in vitro* assays. These models are useful for prioritizing chemicals for further testing. Also, they can guide the choice of the follow-up assay: a chemical triggering an alert in mutagenesis and not in chromosomal damage will be tested, for example, in the Ames test rather than both Ames test and the micronucleus assay.

An example of the so-called integrated testing strategy combining *in silico* and *in vitro* approaches is illustrated by the work of Gubbels-van Hal *et al.* [130].

Combining *in silico* with *in vitro* data may be a way to move toward animal-free genotoxicity and carcinogenicity testing. Effort is ongoing in the *in silico* as well as the *in vitro* areas. Initiatives such as REACH and 7th amendment of the Cosmetics regulation in EU, HPV (High Production Volume) program in the United States, and the DSL (Domestic Substances List) in Canada are incentives for seeking alternatives to animal testing.

Rather than having a defined framework, a flexible approach with an insight into the mechanistic pathways will be needed in order to generate relevant data. The AOP-based approach is leading toward this direction [131]. Collaborative effort is needed in obtaining not only qualitative but also quantitative information to support decision making.

Abbreviations

ADME	absorption, distribution, metabolism, and excretion
AOP	adverse outcome pathway
CA	chromosomal aberration
DMBA	7, 12-dimethylbenzanthracene
DMN	dimethylnitrosamine
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECVAM	European Centre for the Validation of Alternative Methods
ENU	ethylnitrosourea
FHA	fast halo assay
HCS	high-content screening
HTS	high-throughput screening
ILSI/HESI	International Life Sciences Institute/Health and Environmental Sciences Institutes
ITS	integrated testing strategies
MLA	mouse lymphoma assay
MMNCs	micronuclei in mononucleated cells
MMS	methyl methane sulfonate
MNCs	mononucleated cells
OECD	Organisation for Economic Co-operation and Development
OTM	Olive tail moment
PD	population doubling
QSAR	quantitative structure–activity relationship
REACH	Registration, Evaluation, Authorization, and Restriction of Chemicals
RPD	relative population doubling
SCCS	Scientific Committee for Consumer Safety
SCGE	single-cell gel electrophoresis
TK	thymidine kinase
WoE	weight of evidence

References

1 Scientific Committee on Consumer Safety (SCCS) (2012) The SCCS's Notes of Guidance for the Testing of Cosmetic Substances and Their Safety Evaluation, 8th revision. Available at http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_s_006.pdf (last accessed April 2014).

2 Nohynek, G.J., Antignac, E., Re, T., and Toutain, H. (2010) Safety assessment of personal care products/cosmetics and their ingredients. *Toxicol. Appl. Pharmacol.*, **243** (2), 239–259.

3 European Commission (2013) Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. Available at http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:2009R1223:20130711:en:PDF (last accessed April 2014).

4 European Commission (2009) Directive 76/768/EC. Available at http://ec.europa.eu/consumers/sectors/cosmetics/documents/directive/index_en.htm (last accessed April 2014).

5 https://www.cosmeticseurope.eu/safety-and-science-cosmetics-europe/alternative-methods.html (last accessed April 2014).

6 Adler, S., Basketter, D., Creton, S. *et al.* (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects – 2010. *Arch. Toxicol.*, **85** (5), 367–485.

7 National Academy of Sciences (NAS) (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy Calling for a Change of Paradigm in Toxicity Testing*, National Academies Press.

8 http://www.seurat-1.eu/ (last accessed April 2014).

9 http://www.carcinogenomics.eu/ (last accessed April 2014).

10 http://comics.vitamib.com/ (last accessed April 2014).

11 http://www.epa.gov/ncct/toxcast/ (last accessed April 2014).

12 http://epa.gov/ncct/Tox21/ (last accessed April 2014).

13 http://www.safe.nite.go.jp/english/kasinn/qsar/hess-e.html (last accessed April 2014).

14 http://www.chemscreen.eu/ (last accessed April 2014).

15 http://www.hesiglobal.org/i4a/pages/index.cfm?pageid=3492 (last accessed April 2014).

16 Briggs, K., Cases, M., Heard, D.J., Pastor, M., Pognan, F., Sanz, F., Schwab, C.H., Steger-Hartmann, T., Sutter, A., Watson, D.K., and Wichard, J.D. (2012) Inroads to predict *in vivo* toxicology – an introduction to the eTOX project. *Int. J. Mol. Sci.*, **13** (3), 3820–3846.

17 Gabriel, E., Weinreb, M.T., Kapustina, K., and Jacobson, T.C.E. (2009) *In silico* generation of alternative hypotheses using causal mapping (CMAP). *PLoS One*, **4** (4), e5378.

18 Valerio, L.G. (2012) Application of advanced *in silico* methods for predictive modeling and information integration. *Expert Opin. Drug Metab. Toxicol.*, **8** (4), 395–398.

19 OECD (2004) OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship ((Q)SAR) Models.

20 http://www.qsartoolbox.org/ (last accessed April 2014).

21 Ellison, C.M., Enoch, S.J., and Cronin, M.T. (2011) A review of the use of *in silico* methods to predict the chemistry of molecular initiating events related to drug toxicity. *Expert Opin. Drug Metab. Toxicol.*, **7** (12), 1481–1495.

22 Poongavanam, V., Haider, N., and Ecker, G.F. (2012) Fingerprint-based *in silico* models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg. Med. Chem.*, **20** (18), 5388–5395.

23 Peric-Hassler, L., Stjernschantz, E., Oostenbrink, C., and Geerke, D.P. (2013) CYP 2D6 binding affinity predictions using multiple ligand and protein conformations. *Int. J. Mol. Sci.*, **14** (12), 24514–24530.

24 Klepsch, F., Poongavanam, V., and Ecker, G.F. (2014) Ligand and structure-based

classification models for prediction of P-glycoprotein inhibitors. *J. Chem. Inf. Model.*, **54** (1), 218–229.

25 Shah, I. and Wambaugh, J. (2010) Virtual tissues in toxicology. *J. Toxicol. Environ. Health B Crit. Rev.*, **13** (2–4), 314–328.

26 Wetmore, B.A., Wambaugh, J.F., Ferguson, S.S., Li, L., Clewell, H.J., III, Judson, R.S., Freeman, K., Bao, W., Sochaski, M.A., Chu, T.M., Black, M.B., Heal, E., Allen, B., Andersen, M.E., Wolfinger, R.D., and Thomas, R.S. (2013) Relative impact of incorporating pharmacokinetics on predicting *in vivo* hazard and mode of action from high-throughput *in vitro* toxicity assays. *Toxicol. Sci.*, **132** (2), 327–346.

27 Bowes, J., Brown, A.J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and Whitebread, S. (2012) Reducing safety-related drug attrition: the use of *in vitro* pharmacological profiling. *Nat. Rev. Drug Discov.*, **12**, 909–922.

28 Mayer, M. and Yang, J. (2013) Engineered ion channels as emerging tools for chemical biology. *Acc. Chem. Res.*, **46** (12), 2998–3008.

29 Sipes, N.S., Martin, M.T., Kothiya, P., Reif, D.M., Judson, R.S., Richard, A.M., Houck, K.A., Dix, D.J., Kavlock, R.J., and Knudsen, T.B. (2013) Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. *Chem. Res. Toxicol.*, **26** (6), 878–895.

30 Raghuramana, S., Garcia, A.J., Anderson, T.M., Twedea, V.D., Curticea, J.K., Chasea, K., Ramirez, J.M., Olivera, B.M., and Teicherta, R.W. (2014) Defining modulatory inputs into CNS neuronal subclasses by functional pharmacological profiling. *Proc. Natl. Acad. Sci. USA*, **111** (17), 6449–6454.

31 Poongavanam, V., Haider, N., and Ecker, G.F. (2012) Fingerprint-based *in silico* models for the prediction of P-glycoprotein substrates and inhibitors. *Bioorg. Med. Chem.*, **20** (18), 5388–5395.

32 Laustriat, D., Gide, J., and Peschanski, M. (2010) Human pluripotent stem cells in drug discovery and predictive toxicology. *Biochem. Soc. Trans.*, **38** (4), 1051–1057.

33 Maury, Y., Gauthier, M., Peschanski, M., and Martinat, C. (2012) Human pluripotent stem cells for disease modelling and drug screening. *Bioessays.*, **34** (1), 61–71.

34 Sirenko, O., Cromwell, E.F., Crittenden, C., Wignall, J.A., Wright, F.A., and Rusyn, I. (2013) Assessment of beating parameters in human induced pluripotent stem cells enables quantitative *in vitro* screening for cardiotoxicity. *Toxicol. Appl. Pharmacol.*, **273** (3), 500–507.

35 Piersma, A.H., Bosgra, S., van Duursen, M.B., Hermsen, S.A., Jonker, L.R., Kroese, E.D., van der Linden, S.C., Man, H., Roelofs, M.J., Schulpen, S.H., Schwarz, M., Uibel, F., van Vugt-Lussenburg, B.M., Westerhout, J., Wolterbeek, A.P., and van der Burg, B. (2013) Evaluation of an alternative *in vitro* test battery for detecting reproductive toxicants. *Reprod. Toxicol.*, **38** 53–64.

36 Klaric, M., Winkler, J., Vojnits, K., Meganathan, K., Jagtap, S., Ensenat-Waser, R., Hescheler, J., Sachinidis, A., and Bremer-Hoffmann, S. (2013) Current status of human pluripotent stem cell based *in vitro* toxicity tests. *Front. Biosci.*, **5**, 118–133.

37 Buzanska, L., Zychowicz, M., Ruiz, A., Ceriotti, L., Coecke, S., Rauscher, H., Sobanski, T., Whelan, M., Domanska-Janik, K., Colpo, P., and Rossi, F. (2010) Neural stem cells from human cord blood on bioengineered surfaces – novel approach to multiparameter bio-tests. *Toxicology*, **270** (1), 35–42.

38 Minowa, Y., Kondo, C., Uehara, T., Morikawa, Y., Okuno, Y., Nakatsu, N., Ono, A., Maruyama, T., Kato, I., Yamate, J., Yamada, H., Ohno, Y., and Urushidani, T. (2012) Toxicogenomic multigene biomarker for predicting the future onset of proximal tubular injury in rats. *Toxicology*, **297** (1–3), 47–56.

39 Doktorova, T.Y., Ellinger-Ziegelbauer, H., Vinken, M., Vanhaecke, T., van Delft, J., Kleinjans, J., Ahr, H.J., and Rogiers, V. (2012) Comparison of hepatocarcinogen-induced gene expression profiles in conventional primary rat hepatocytes

with *in vivo* rat liver. *Arch. Toxicol.*, **86** (9), 1399–1411.
40 Currie, R.A. (2012) Toxicogenomics: the challenges and opportunities to identify biomarkers, signatures and thresholds to support mode-of-action. *Mutat. Res.*, **746** (2), 97–103.
41 Rusyn, I., Sedykh, A., Low, Y., Guyton, K. Z., and Tropsha, A. (2012) Predictive modeling of chemical hazard by integrating numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol. Sci.*, **127** (1), 1–9.
42 Van Summeren, A., Renes, J., van Delft, J.H., Kleinjans, J.C., and Mariman, E.C. (2012) Proteomics in the search for mechanisms and biomarkers of drug-induced hepatotoxicity. *Toxicol. In Vitro*, **26** (3), 373–385.
43 Bouhifd, M., Hartung, T., Hogberg, H.T., Kleensang, A., and Zhao, L. (2013) Review: toxicometabolomics. *J. Appl. Toxicol.*, **33** (12), 1365–1383.
44 Ramirez, T., Daneshian, M., Kamp, H., Bois, F.Y., Clench, M.R., Coen, M., Donley, B., Fischer, S.M., Ekman, D.R., Fabian, E., Guillou, C., Heuer, J., Hogberg, H.T., Jungnickel, H., Keun, H. C., Krennrich, G., Krupp, E., Luch, A., Noor, F., Peter, E., Riefke, B., Seymour, M., Skinner, N., Smirnova, L., Verheij, E., Wagner, S., Hartung, T., van Ravenzwaay, B., and Leist, M. (2013) Metabolomics in toxicology and preclinical research. *ALTEX*, **30** (2), 209–225.
45 Hirobe, T., Hasegawa, K., Furuya, R., Fujiwara, R., and Sato, K. (2013) Effects of fibroblast-derived factors on the proliferation and differentiation of human melanocytes in culture. *J. Dermatol. Sci.*, **71** (1), 45–57.
46 Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H.J. (2008) Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term *in vivo* studies. *Mutat. Res.*, **637** (1–2), 23–39.
47 Sarrif, A., van Delft, J.H., van Schooten, F.J., Gant, T.W., Elliott, B.M., van Ravenzwaay, B., van Steeg, H., and Vrijhof, H. (2005) Toxicogenomics in genetic toxicology and hazard determination: introduction and overview. *Mutat. Res.*, **575** (1–2), 1–3.
48 Sarrif, A., van Delft, J.H., Gant, T.W., Kleinjans, J.C., and van Vliet, E. (2005) Toxicogenomics in genetic toxicology and hazard determination – concluding remarks. *Mutat. Res.*, **575** (1–2), 116–117.
49 Thybaud, V., Le Fevre, A.C., and Boitier, E. (2007) Application of toxicogenomics to genetic toxicology risk assessment. *Environ. Mol. Mutagen.*, **48**, 369–379.
50 Diaz, D., Scott, A., Carmichael, P., Shi, W., and Costales, C. (2007) Evaluation of an automated *in vitro* micronucleus assay in CHO-K1 cells. *Mutat. Res.*, **630** (1–2), 1–13.
51 Kalusche, G., Smith, L., and Thomas, N. (2005) Ultra High Speed Automated *In-Vitro* Micronucleus Analysis. SBS Congress UK, Session: ADME/Tox – Poster Session (Presentation: P05007).
52 Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W., and Kavlock, R.J. (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol. Sci.*, **95** (1), 5–12.
53 http://www.ncats.nih.gov/research/reengineering/tissue-chip/tissue-chip.html (last accessed April 2014).
54 Alcendor, D.J., Block, F.E., 3rd, Cliffel, D. E., Daniels, J.S., Ellacott, K.L., Goodwin, C.R., Hofmeister, L.H., Li, D., Markov, D. A., May, J.C., McCawley, L.J., McLaughlin, B., McLean, J.A., Niswender, K.D., Pensabene, V., Seale, K.T., Sherrod, S.D., Sung, H.J., Tabb, D.L., Webb, D.J., and Wikswo, J.P. (2013) Neurovascular unit on a chip: implications for translational applications. *Stem. Cell Res. Ther.*, **4** (Suppl. 1), S18.
55 Mueller, D., Krämer, L., Hoffmann, E., Klein, S., and Noor, F. (2014) 3D organotypic HepaRG cultures as *in vitro* model for acute and repeated dose toxicity studies. *Toxicol. In Vitro*, **28** (1), 104–112.
56 Soldatow, V.Y., Lecluyse, E.L., Griffith, L. G., and Rusyn, I. (2013) *In vitro* models for liver toxicity testing. *Toxicol. Res.*, **2** (1), 23–39.

57 Gunness, P., Mueller, D., Shevchenko, V., Heinzle, E., Ingelman-Sundberg, M., and Noor, F. (2013) 3D organotypic cultures of human HepaRG cells: a tool for *in vitro* toxicity studies. *Toxicol. Sci.*, **133** (1), 67–78.

58 Dash, A., Blackman, B.R., and Wamhoff, B.R. (2012) Organotypic systems in drug metabolism and toxicity: challenges and opportunities. *Expert Opin. Drug Metab. Toxicol.*, **8** (8), 999–1014.

59 LeCluyse, E.L., Witek, R.P., Andersen, M. E., and Powers, M.J. (2012) Organotypic liver culture models: meeting current challenges in toxicity testing. *Crit. Rev. Toxicol.*, **42** (6), 501–548.

60 Pappinen, S., Pryazhnikov, E., Khiroug, L., Ericson, M.B., Yliperttula, M., and Urtti, A. (2012) Organotypic cell cultures and two-photon imaging: tools for *in vitro* and *in vivo* assessment of percutaneous drug delivery and skin toxicity. *J. Control. Release*, **161** (2), 656–667.

61 Kidambi, S., Yarmush, R.S., Novik, E., Chao, P., Yarmush, M.L., and Nahmias, Y. (2009) Oxygen-mediated enhancement of primary hepatocyte metabolism, functional polarization, gene expression, and drug clearance. *Proc. Natl. Acad. Sci. USA*, **106** (37), 15714–15719.

62 Maguire, T.J., Novik, E., Chao, P., Barminko, J., Nahmias, Y., Yarmush, M.L., and Cheng, K.C. (2009) Design and application of microfluidic systems for *in vitro* pharmacokinetic evaluation of drug candidates. *Curr. Drug Metab.*, **10** (10), 1192–1199.

63 Thomas, R.S., Philbert, M.A., Auerbach, S.S., Wetmore, B.A., Devito, M.J., Cote, I., Rowlands, J.C., Whelan, M.P., Hays, S.M., Andersen, M.E., Meek, M.E., Reiter, L.W., Lambert, J.C., Clewell, H.J., 3rd, Stephens, M.L., Zhao, Q.J., Wesselkamper, S.C., Flowers, L., Carney, E.W., Pastoor, T.P., Petersen, D.D., Yauk, C.L., and Nong, A. (2013) Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. *Toxicol. Sci.*, **136** (1), 4–18.

64 Yoon, M., Efremenko, A., Blaauboer, B.J., and Clewell, H.J. (2014) Evaluation of simple *in vitro* to *in vivo* extrapolation approaches for environmental compounds. *Toxicol. In Vitro*, **28** (2), 164–170.

65 Kunze, A., Huwyler, J., Poller, B., Gutmann, H., and Camenisch, G. (2014) *In vitro–in vivo* extrapolation method to predict human renal clearance of drugs. *J. Pharm. Sci.*, **103** (3), 994–1001.

66 Kirkland, D., Aardema, M., Henderson, L., and Müller, L. (2005) Evaluation of the ability of a battery of three *in vitro* genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. *Mutat. Res.*, **584** (1–2), 1–256. Erratum: *Mutat. Res.*, 2005, **588** (1), 70.

67 Fowler, P., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhler, S., and Carmichael, P. (2012) Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. I. Choice of cell type. *Mutat. Res.*, **742** (1–2), 11–25.

68 Fowler, P., Smith, R., Smith, K., Young, J., Jeffrey, L., Carmichael, P., Kirkland, D., and Pfuhler, S. (2014) Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. III. Sensitivity of human cell types to known genotoxic agents. *Mutat. Res.* doi: 10.1016/j.mrgentox.2014.03.001.

69 Fowler, P., Smith, R., Smith, K., Young, J., Jeffrey, L., Kirkland, D., Pfuhler, S., and Carmichael, P. (2012) Reduction of misleading ("false") positive results in mammalian cell genotoxicity assays. II. Importance of accurate toxicity measurement. *Mutat. Res.*, **747** (1), 104–117.

70 Galloway, S., Lorge, E., Aardema, M.J., Eastmond, D., Fellows, M., Heflich, R., Kirkland, D., Levy, D.D., Lynch, A.M., Marzin, D., Morita, T., Schuler, M., and Speit, G. (2011) Workshop summary: top concentration for *in vitro* mammalian cell genotoxicity assays; and report from working group on toxicity measures and top concentration for *in vitro* cytogenetics assays (chromosome aberrations and micronucleus). *Mutat. Res.*, **723** (2), 77–83.

71 Curren, R.D., Mun, G.C., Gibson, D.P., and Aardema, M.J. (2006) Development

of a method for assessing micronucleus induction in a 3D human skin model (EpiDerm). *Mutat. Res.*, **607** (2), 192–204.

72 Flamand, N., Marrot, L., Belaidi, J.P., Bourouf, L., Dourille, E., Feltes, M., and Meunier, J.R. (2006) Development of genotoxicity test procedures with Episkin, a reconstructed human skin model: towards new tools for *in vitro* risk assessment of dermally applied compounds? *Mutat. Res.*, **606**, 39–51.

73 Hu, T., Kaluzhny, Y., Mun, G.C., Barnett, B., Karetsky, V., Wilt, N., Klausner, M., Curren, R.D., and Aardema, M.J. (2009) Intralaboratory and interlaboratory evaluation of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat. Res.*, **673** (2), 100–108.

74 Aardema, M.J., Barnett, B.C., Khambatta, Z., Reisinger, K., Ouedraogo-Arras, G., Faquet, B., Ginestet, A.C., Mun, G.C., Dahl, E.L., Hewitt, N.J., Corvi, R., and Curren, R.D. (2010) International prevalidation studies of the EpiDerm 3D human reconstructed skin micronucleus (RSMN) assay: transferability and reproducibility. *Mutat. Res.*, **701** (2), 123–131.

75 Reus, A.A., Reisinger, K., Downs, T.R., Carr, G.J., Zeller, A., Corvi, R., Krul, C.A., and Pfuhler, S. (2013) Comet assay in reconstructed 3D human epidermal skin models – investigation of intra- and inter-laboratory reproducibility with coded chemicals. *Mutagenesis*, **28** (6), 709–720.

76 Fenech, M. and Morley, A.A. (1985) The effect of donor age on spontaneous and induced micronuclei. *Mutat. Res.*, **148** (1–2), 99–105.

77 Singh, N.P., McCoy, M.T., Tice, R.R., and Schneider, E.L. (1988) A simple technique for quantitation of low levels of DNA damage in individual cells. *Exp. Cell Res.*, **175**, 184–191.

78 Olive, P.L., Banath, J.P., and Durand, R.E. (1990) Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells measured using the "comet" assay. *Radiat. Res.*, **122**, 86–94.

79 Ostling, O. and Johanson, K.J. (1984) Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem. Biophys. Res. Commun.*, **123**, 291–298.

80 Sestili, P., Martinelli, C., and Stocchi, V. (2006) The fast halo assay: an improved method to quantify genomic DNA strand breakage at the single-cell level. *Mutat. Res.*, **607**, 205–214.

81 Kirkland, D., Pfuhler, S., Tweats, D., Aardema, M., Corvi, R., Darroudi, F., Elhajouji, A., Glatt, H., Hastwell, P., Hayashi, M., Kasper, P., Kirchner, S., Lynch, A., Marzin, D., Maurici, D., Meunier, J.R., Muller, L., Nohynek, G., Parry, J., Parry, E., Thybaud, V., Tice, R., van Benthem, J., Vanparys, P., and White, P. (2007) How to reduce false positive results when undertaking *in vitro* genotoxicity testing and thus avoid unnecessary follow-up animal tests: report of an ECVAM Workshop. *Mutat. Res.*, **628**, 31–55.

82 Platel, A., Nesslany, F., Gervais, V., and Marzin, D. (2009) Study of oxidative DNA damage in TK6 human lymphoblastoid cells by use of the *in vitro* micronucleus test: determination of no-observed-effect levels. *Mutat. Res.*, **678**, 30–37.

83 Kirkland, D., Kasper, P., Muller, L., Corvi, R., and Speit, G. (2008) Recommended lists of genotoxic and non-genotoxic chemicals for assessment of the performance of new or improved genotoxicity tests: a follow-up to an ECVAM workshop. *Mutat. Res.*, **653**, 99–108.

84 NTP, http://ntp.niehs.nih.gov/index.cfm (last accessed April 2014).

85 Zeiger, E., Ashby, J., Bakale, G., Enslein, K., Klopman, G., and Rosenkranz, H.S. (1996) Prediction of *Salmonella* mutagenicity. *Mutagenesis*, **11** (5), 471–484.

86 Matsushima, T., Hayashi, M., Matsuoka, A., Ishidate, M., Jr. Miura, K.F., Shimizu, H., Suzuki, Y., Morimoto, K., Ogura, H., Mure, K., Koshi, K., and Sofuni, T. (1999) Validation study of the *in vitro* micronucleus test in a Chinese hamster

lung cell line (CHL/IU). *Mutagenesis*, **14** (6), 569–580.

87 Ishidate, M.J., Harnois, M.C., and Sofuni, T. (1988) A comparative analysis of data on the clastogenicity of 951 chemical substances tested in mammalian cell cultures. *Mutat. Res.*, **195**, 151–213.

88 Soukup, S.W. and Au, W. (1975) The effect of ethylnitrosourea on chromosome aberrations *in vitro* and *in vivo*. *Humangenetik*, **29** (4), 319–328.

89 Mavournin, K.H., Blakey, D.H., Cimino, M.C., Salamone, M.F., and Heddle, J.A. (1990) The *in vivo* micronucleus assay in mammalian bone marrow and peripheral blood. A report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat. Res.*, **239** (1), 29–80.

90 Nesnow, S., Gold, A., Mohapatra, N., Sangaiah, R., Bryant, B.J., Rudo, K., MacNair, P., and Ellis, S. (1986) Metabolic activation pathways of cyclopenta-fused PAH and their relationship to genetic and carcinogenic activity. *Prog. Clin. Biol. Res.*, **209A**, 515–522.

91 Mitchell, A.D., Rudd, C.J., and Caspary, W.J. (1988) Evaluation of the L5178Y mouse lymphoma cell mutagenesis assay: intralaboratory results for 63 coded chemicals tested at SRI International. *Environ. Mol. Mutagen.*, **12** (13), 37–102.

92 Lambert, I.B., Singer, T.M., Boucher, S.E., and Douglas, G.R. (2005) Detailed review of transgenic rodent mutation assays. *Mutat. Res.*, **590** (1–3) 1–280.

93 Mitchell, A.D., Auletta, A.E., Clive, D., Kirby, P.E., Moore, M.M., and Myhr, B.C. (1997) The L5178Y/tk$^{+/-}$ mouse lymphoma specific gene and chromosomal mutation assay: a phase III report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat. Res.*, **394**, 177–303.

94 Otteneder, M. and Lutz, W.K. (1999) Correlation of DNA adduct levels with tumor incidence: carcinogenic potency of DNA adducts. *Mutat. Res.*, **424** (1–2) 237–247.

95 Gold, L.S., Slone, T.H., Manley, N.B., Garfinkel, G.B., Hudes, E.S., Rohrbach, L., and Ames, B.N. (1991) The Carcinogenic Potency Database: analyses of 4000 chronic animal cancer experiments published in the general literature and by the U.S. National Cancer Institute/ National Toxicology Program. *Environ. Health Perspect.*, **96**, 11–15.

96 Morita, T., Asano, N., Awogi, T., Sasaki, Y.F., Sato, S., Shimada, H., Sutou, S., Suzuki, T., Wakata, A., Sofuni, T., and Hayashi, M. (1997) Evaluation of the rodent micronucleus assay in the screening of IARC carcinogens (groups 1, 2A and 2B): the summary report of the 6th collaborative study by CSGMT/JEMS MMS. Collaborative Study of the Micronucleus Group Test. Mammalian Mutagenicity Study Group. *Mutat. Res.*, **389** (1), 3–122. Erratum: *Mutat. Res.*, 1997, **391** (3), 259–267.

97 Madle, E., Korte, A., and Beek, B. (1986) Species differences in mutagenicity testing. I. Micronucleus and SCE tests in rats, mice, and Chinese hamsters with aflatoxin B1. *Teratog. Carcinog. Mutagen.*, **6** (1), 1–13.

98 Sinsheimer, J.E., Hooberman, B.H., Das, S.K., Brezzell, M.D., and You, Z. (1992) The *in vivo* and *in vitro* genotoxicity of aromatic amines in relationship to the genotoxicity of benzidine. *Mutat. Res.*, **268** (2), 255–264.

99 Slesinski, R.S., Guzzie, P.J., Putman, D.L., and Ballantyne, B. (1988) *In vitro* and *in vivo* evaluation of the genotoxic potential of 2-ethyl-1,3-hexanediol. *Toxicology*, **53** (2–3), 179–198.

100 Gold, L.S. (2004) The Carcinogenic Potency Project. Available at http://potency.berkeley.edu/cpdb.html (last accessed April 2014).

101 Mortelmans, K., Haworth, S., Lawlor, T., Speck, W., Tainer, B., and Zeiger, E. (1986) *Salmonella* mutagenicity tests. II. Results from the testing of 270 chemicals. *Environ. Mutagen.*, **8** (7), 1–119.

102 Stemp, G., Pascoe, S., and Gatehouse, D. (1989) *In vitro* and *in vivo* cytogenetic studies of three beta-lactam antibiotics (penicillin VK, ampicillin and carbenicillin). *Mutagenesis*, **4** (6), 439–445.

103 Gulati, D.K., Witt, K., Anderson, B., Zeiger, E., and Shelby, M.D. (1989) Chromosome aberration and sister

chromatid exchange tests in Chinese hamster ovary cells *in vitro*. III. Results with 27 chemicals. *Environ. Mol. Mutagen.*, **13** (2), 133–193.

104 Shelby, M.D. and Witt, K.L. (1995) Comparison of results from mouse bone marrow chromosome aberration and micronucleus tests. *Environ. Mol. Mutagen.*, **25** (4), 302–313.

105 IARC (1983) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Available at http://monographs.iarc.fr/ENG/Monographs/vol1-42/mono32.pdf (last accessed April 2014).

106 Hilliard, C.A., Armstrong, M.J., Bradt, C. I., Hill, R.B., Greenwood, S.K., and Galloway, S.M. (1998) Chromosome aberrations *in vitro* related to cytotoxicity of nonmutagenic chemicals and metabolic poisons. *Environ. Mol. Mutagen.*, **31**, 316–326.

107 Zeiger, E., Anderson, B., Haworth, S., Lawlor, T., and Mortelmans, K. (1988) *Salmonella* mutagenicity tests. IV. Results from the testing of 300 chemicals. *Environ. Mol. Mutagen.*, **11** (12), 1–157.

108 UNEP Chemicals;1; (2004) Screening Information Data Set (SIDS) for High Volume Chemicals. Available at http://www.chem.unep.ch/irptc/sids/oecdsids/sidspub.html (last accessed April 2014).

109 Oshiro, Y., Piper, C.E., Balwierz, P.S., and Soelter, S.G. (1991) Chinese hamster ovary cell assays for mutation and chromosome damage: data from non-carcinogens. *J. Appl. Toxicol.*, **11**, 167–177.

110 Shelby, M.D. and Stasiewicz, S. (1984) Chemicals showing no evidence of carcinogenicity in long-term, two-species rodent studies: the need for short-term test data. *Environ. Mutagen.*, **6**, 871–878.

111 Zeiger, E., Anderson, B., Haworth, S., Lawlor, T., Mortelmans, K., and Speck, W. (1987) *Salmonella* mutagenicity tests. III. Results from the testing of 225 chemicals. *Environ. Mutagen.*, **9** (9), 1–109.

112 McFee, A.F., Jauhar, P.P., Lowe, K.W., MacGregor, J.T., and Wehr, C.M. (1989) Assays of three carcinogen/non-carcinogen chemical pairs for *in vivo* induction of chromosome aberrations, sister chromatid exchanges and micronuclei. *Environ. Mol. Mutagen.*, **14** (4), 207–220.

113 Shelby, M.D., Erexson, G.L., Hook, G.J., and Tice, R.R. (1993) Evaluation of a three-exposure mouse bone marrow micronucleus protocol: results with 49 chemicals. *Environ. Mol. Mutagen.*, **21**, 160–179.

114 Ivett, J.L., Brown, B.M., Rodgers, C., Anderson, B.E., Resnick, M.A., and Zeiger, E. (1989) Chromosomal aberration and sister chromatid exchange tests in Chinese hamster ovary cells *in vitro*. IV. Results for l5 chemicals. *Environ. Mol. Mutagen.*, **14**, 165–187.

115 Meintieres, S., Biola, A., Pallardy, M., and Marzin, D. (2003) Using CTLL-2 and CTLL-2 bcl2 cells to avoid interference by apoptosis in the *in vitro* micronucleus test. *Environ. Mol. Mutagen.*, **41**, 14–27.

116 Duvoix, A., Blasius, R., Delhalle, S., Schnekenburger, M., Morceau, F., Henry, E., Dicato, M., and Diederich, M. (2005) Chemopreventive and therapeutic effects of curcumin. *Cancer Lett.*, **223** (2), 181–190.

117 Borenfreund, E., Babich, H., and Martin-Alguacil, N. (1988) Comparisons of two *in vitro* cytotoxicity assays – the neutral red (NR) and tetrazolium MTT tests. *Toxicol. In Vitro*, **2**, 1–6.

118 Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, **65**, 55–63.

119 Fenech, M. (2006) Cytokinesis-block micronucleus assay evolves into a "cytome" assay of chromosomal instability, mitotic dysfunction and cell death. *Mutat. Res.*, **600**, 58–66.

120 Fenech, M. (2000) The *in vitro* micronucleus technique. *Mutat. Res.*, **455**, 81–95.

121 Fairbairn, D.W., Walburger, D.K., Fairbairn, J.J., and O'Neill, K.L. (1996) Key morphologic changes and DNA strand breaks in human lymphoid cells: discriminating apoptosis from necrosis. *Scanning*, **18**, 407–416.

122 Kizilian, N., Wilkins, R.C., Reinhardt, P., Ferrarotto, C., McLean, J.R., and McNamee, J.P. (1999) Silver-stained comet assay for detection of apoptosis. *Biotechniques*, **27**, 926–930.

123 Dahl, E.L., Curren, R., Barnett, B.C., Khambatta, Z., Reisinger, K., Ouedraogo, G., Faquet, B., Ginestet, A.C., Mun, G., Hewitt, N.J., Carr, G., Pfuhler, S., and Aardema, M.J. (2011) The reconstructed skin micronucleus assay (RSMN) in EpiDerm^{TM}: detailed protocol and harmonized scoring atlas. *Mutat. Res.*, **720** (1–2), 42–52.

124 Hewitt, N.J., Edwards, R.J., Fritsche, E., Goebel, C., Aeby, P., Scheel, J., Reisinger, K., Ouédraogo, G., Duche, D., Eilstein, J., Latil, A., Kenny, J., Moore, C., Kuehnl, J., Barroso, J., Fautz, R., and Pfuhler, S. (2013) Use of human *in vitro* skin models for accurate and ethical risk assessment: metabolic considerations. *Toxicol. Sci.*, **133** (2), 209–217.

125 Aardema, M.J., Barnett, B.B., Mun, G.C., Dahl, E.L., Curren, R.D., Hewitt, N.J., and Pfuhler, S. (2013) Evaluation of chemicals requiring metabolic activation in the EpiDerm^{TM} 3D human reconstructed skin micronucleus (RSMN) assay. *Mutat. Res.*, **750** (1–2), 40–49.

126 Pfuhler, S., Fautz, R., Ouedraogo, G., Latil, A., Kenny, J., Moore, C., Diembeck, W., Hewitt, N.J., Reisinger, K., and Barroso, J. (2014) The Cosmetics Europe strategy for animal-free genotoxicity testing: project status up-date. *Toxicol. In Vitro*, **28** (1), 18–23.

127 Dearfield, K.L., Thybaud, V., Cimino, M.C., Custer, L., Czich, A., Harvey, J.S., Hester, S., Kim, J.H., Kirkland, D., Levy, D.D., Lorge, E., Moore, M.M., Ouédraogo-Arras, G., Schuler, M., Suter, W., Sweder, K., Tarlo, K., van Benthem, J., van Goethem, F., and Witt, K.L. (2011) Follow-up actions from positive results of *in vitro* genetic toxicity testing. *Environ. Mol. Mutagen.*, **52** (3), 177–204.

128 National Research Council (2007) *Toxicity in the 21st Century: A Vision and a Strategy*, National Academies Press.

129 Adeleye, Y., Andersen, M., Clewell, R., Davies, M., Dent, M., Edwards, S., Fowler, P., Malcomber, S., Nicol, B., Scott, A., Scott, S., Sun, B., Westmoreland, C., White, A., Zhang, Q., and Carmichael, P.L. (2014) Implementing Toxicity Testing in the 21st Century (TT21C): making safety decisions using toxicity pathways, and progress in a prototype risk assessment. *Toxicology*, doi: 10.1016/j.tox.2014.02.007.

130 Gubbels-van Hal, W.M., Blaauboer, B.J., Barentsen, H.M., Hoitink, M.A., Meerts, I.A., and van der Hoeven, J.C. (2005) An alternative approach for the safety evaluation of new and existing chemicals, an exercise in integrated testing. *Regul. Toxicol. Pharmacol.*, **42**, 284–295.

131 Landesmann, B., Mennecozzi, M., Berggren, E., and Whelan, M. (2013) Adverse outcome pathway-based screening strategies for an animal-free safety assessment of chemicals. *Altern. Lab. Anim.*, **41** (6), 461–471.

13
Using Pluripotent Stem Cells and Their Progeny as an *In Vitro* Model to Assess (Developmental) Neurotoxicity

Lisa Hoelting, Marcel Leist, and Luc Stoppini

13.1
Introduction

Embryonic stem cells (ESCs) are self-renewing pluripotent cells derived from the inner cell mass (ICM) of blastocyst of the developing embryo [1]. *In vitro*, ESCs have an indefinite proliferation potential while maintaining the capability to differentiate into theoretically any cell type in the body [2]. Neural stem cells (NSCs) not only exist in the developing embryo but are also present in specific regions of the adult human brain. Once isolated, adult NSCs have a more limited proliferation potential and their differentiation capacity is restricted to neurons, astrocytes, and oligodendrocytes [3,4].

During the past 30 years, stem cells (SCs) have become a promising research tool for disease modeling and regenerative medicine. Over the years, much effort has been spent on the generation of disease-pertinent somatic cells to investigate *in vitro* molecular mechanism of disease-specific pathophysiologies and their progression. This may lead to a better understanding of the disease and offers new therapies [5–7]. Further important progress was obtained by the newly achieved success in generation of induced pluripotent stem cells (iPSCs). Overexpression of a set of transcription factors (OCT-4, SOX2, KLF4, and v-myc) can reprogram somatic cells into pluripotent stem cells [8]. The new technology enables the production of patient-specific iPSCs to study diseases such as amyotrophic lateral sclerosis, familial dysautonomia, or Huntington's disease [9–11]. Tremendous improvements have been made in quality and efficiency of *in vitro* differentiation and numerous protocols have been established to differentiate human ESCs or patient-specific iPSCs into clinically relevant somatic cell types such as cardiomyocytes [10,12,13], hematopoietic cells [14], megakaryocytes [15], insulin-producing cells [16–19], and neurons [20–22]. However, the aim to use these cells to treat diseases such as diabetes mellitus, Parkinson's disease, and Alzheimer's disease by transplanting *in vitro*-generated human SC-based cells into the patient and thereby replacing dysfunctional cells is still challenging and many difficulties have to be overcome to routinely use SCs for safe therapies in the future [23–25]. Currently running clinical trials on the use of retinal pigment

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

epithelial cells to treat dry macular degeneration are, however, promising and the reported side effects are only minor [26].

In contrast, much more success is expected in the short run by using stem cells for pharmacological and toxicological purposes [27]. Only a minority (~2%) of publications on ESCs focus on the use of stem cells for drug and toxicity screenings, and high-impact journals tend not to publish such work. Nevertheless, researchers have gained much knowledge from the field of disease modeling and biomedical research [28,29]. For instance, SCs have proved to be one of the most important tools to assess (developmental) neurotoxicity [30,31].

Neurotoxicity (NT) is defined as adverse effects of industrial and environmental chemicals, pharmaceutical drugs, food ingredients, and cosmetics on the nervous system during development and adulthood [32]. The nervous system is a very complex organ that can be anatomically divided into the central nervous system consisting of brain and spinal cord and the peripheral nervous system comprising afferent and efferent nerves. The peripheral nervous system is functionally divided into motor, sensory, and autonomic functions; the central nervous system serves many more functions, and therefore many different types of neurotoxicity can be observed. For instance, degeneration of peripheral neurons may result in disruptions of motor or sensory functions and central nervous effects may manifest chronically by reduced cognitive function or attention span, or acutely by dizziness, nausea, and disorientation [33]. In particular, *in utero* or postpartum exposures to toxic compounds during sensitive stages of the development of the nervous system can lead to impaired nervous system function, even long time after removal of the original stimulus [34–36]. In the United States, one of six children is being born with neurodevelopmental disorders such as attention deficit disorders, reduced intellect, mental retardation, autism, and cognitive and behavior alterations [37]. Over the past few decades, epidemiological studies and clinical evidence have revealed a link between neurodevelopmental disorders and exposure to chemicals during early fetal development [38,39]. The developing brain is more vulnerable to hazardous influences than adult brain. Since it is impossible to predict adverse effects of compounds on the developing nervous system based on their direct neurotoxicity potential, special attention is paid to the use of developmental neurotoxicity (DNT) studies in human health risk assessment [40,41]. In 2007, the Organisation for Economic Co-operation and Development (OECD) adopted the developmental neurotoxicity guideline (TG426) to assess environmental and industrial chemicals for their potential to cause DNT [42,43]. In 2009, a review pointed out that merely 100 compounds have been tested using the DNT guideline. Most of them were pesticides and only eight industrial chemicals have been examined [40]. Another survey investigated the DNT potential of 174 compounds by using neurobehavioral risk assessment [44].

Currently used methods to analyze DNT are based on animals, which were exposed to chemicals during critical prenatal and early postnatal periods of the nervous system development. Offspring was then tested under the DNT guideline, containing functional, behavioral, and neuropathological analyses of

the nervous system. However, evaluation of gross neurological and behavioral parameters such as physical development, behavioral ontogeny, motor activity, learning, and memory and postmortem evaluation of brain weights were time and cost intensive in terms of animal use, laboratory equipment, and materials. Thus, due to high cost of animal studies (3 billion dollars per year) and low throughput, the DNT capability of the majority of compounds is still undetermined [31,45,46].

In the field of toxicity assessment, neurotoxicity and in particular DNT is one of the most challenging disciplines [47,48]. Adverse effects on neurodevelopment could, among others, result in altered neuronal cell population in the absence of cell death and thus may not be detected by routine screens. Moreover, there are differences in the brain development of humans and animals. For instance, the human nervous system development at the end of the second trimester of gestation correlates to the brain developmental stage of a newborn rat [49]. Moreover, toxicity studies dealing with the question how one species models for another have shown that rats and rabbits predict each other in only 60% of the tested chemicals. It can be assumed that the correlation among the different species is higher than the correlation between these animals and humans [50,51]. Thus, uncertainty factors have to be included into human health risk assessment based on animal data. Experiences in the past in the field of drug design have shown that a large number of new drug candidates failed in clinical trials despite extensive testing in animal studies [52], and a sizable number of already marketed drugs had to be withdrawn due to late-appearing side effects [53].

In order to improve the quality and the efficiency of human DNT risk assessment, a paradigm shift in the field of toxicology took place. Great effort is spent on the understanding of cellular and molecular mechanisms that trigger neurodevelopment impairment. According to the National Research Council report of its vision of toxicity testing in the twenty-first century (Tox21), knowledge on basic cellular and molecular toxicity mechanisms should be used to develop a new *in vitro* test system [54,55]. In particular, in the field of (developmental) neurotoxicity, *in vitro* test systems should be modeling crucial events of the *in vivo* neurodevelopment, containing endpoints representing mechanisms of (developmental) neurotoxicity or specificity of a nervous system response [31]. Human stem cells are the most promising source of human cells that fulfill these criteria and are therefore a promising tool to assess human-specific (developmental) neurotoxicity [56–59].

In this chapter, we will focus on the use of (human) pluripotent stem cells (PSCs) and their progeny to assess (developmental) neurotoxicity.

13.2 Neurodevelopment *In Vivo*

In humans, the nervous system develops from a single cell layer during early embryogenesis into a complex functional neuronal circuit in young adults.

Inhibition of bone morphogenetic protein (BMP) signaling by the neural inducing factors noggin, follistatin, and chordin induces neurodevelopment in a part of the ectodermal layer [60–62]. In combination with the activation of the fibroblast growth factor (FGF), WNT, and insulin-like growth factor (IGF) signaling [63,64], the neural plate is formed and develops further into the neural groove that folds up into the neural tube at approximately week 4 postfertilization [65,66].

The neural tube is formed by neuroepithelial progenitor (NEP) cells. These primary neural stem cells are the origin of nearly all neurons and glial cells of the brain and spinal cord. The newborn neurons have to migrate from their place of origin to their terminal destination, undergo neuritogenesis and synaptogenesis, and assemble a functional circuit to communicate with both nearby and distant cells [67–70]. The process of neurulation is highly controlled and has to take place within a critical time window during embryogenesis. Thus, neural tube closure is very susceptible to even small disturbances that may result in a wide spectrum of morphological malformations [71]. The most common neural tube defects are spina bifida and anencephaly [72]. Although more than 200 genes have been identified that are associated with neural tube defects in animal models, the mechanism by which these genes in combination with environmental factors affect neural tube closure is still unknown [73]. Drugs may disturb this complex signaling. For instance, multiple prospective and retrospective studies have demonstrated that the consumption of the antiepileptic drug valproic acid (VPA) during pregnancy is associated with a significantly increased incidence of neural tube defects [74].

During neural tube closure, another neural stem cell class arises. At the lateral border of the neural plate, neural crest stem cells (NCSCs) undergo an epithelial-to-mesenchymal transition (EMT). They delaminate from the neuroepithelium and start to migrate along stereotypical pathways. NCSCs contribute to a variety of different cell types of neuronal and non-neuronal origin. For example, NCSCs give rise to cells of the peripheral nervous system, melanocytes, smooth muscle cells, and cells forming bone and cartilage [75,76]. Impairment of key developmental processes during neural crest development, such as differential cell proliferation, apoptosis, migration, and differentiation, is associated with a wide range of birth defects [77,78]. For instance, disturbance of vagal NCSC migration may result in Hirschsprung's disease, which is characterized by a congenital malformation of the hindgut due to the absence of parasympathetic intrinsic ganglion cells in the submucosal and myenteric plexuses [79,80]. Besides genetic defects, also xenobiotics can affect NCSCs. Ethanol or the pesticide triadimefon are among these compounds [81,82].

Due to inaccessibility of the human embryo, studies of early human development are limited. Detailed knowledge of human neurodevelopment is not available, or it is only inferred from studies performed on animals. Even though the human nervous system differs from that of model organisms, in terms of size, complexity, folding, and period of development, there are also similarities with respect to the highly orchestrated neurodevelopment. Uncommitted progenitor

cells have to proliferate and differentiate in a temporally and spatially restricted sequence to produce specific neuronal and glial subtypes in defined quantities at stereotyped positions within the nervous system. One fundamental principle of neurodevelopment is a progressing differentiation from multipotent neural stem cells with a broad development potential to distinct lineage-committed cell types, which are highly specialized. The basic mechanism coordinating these complex developmental programs relies on time-dependent exposure to gradients of different types and concentrations of growth factors, vitamins, metabolites, or soluble or tethered ligands. Within each cell, a variety of signals are processed, and they activate cognate cascades of transcription factors, which further trigger defined cellular changes for specific cell fate determinations [83,84]. Deregulation of these precise processes is associated with several neurodevelopmental disorders. For example, holoprosencephaly and other congenital malformations could be associated with altered sonic hedgehog (shh) signaling during development [85–87]. Moreover, there is strong evidence that diseases such as schizophrenia or autism are caused by a combination of genetic and environmental factors that trigger early neurodevelopment impairment [88,89].

Neurodevelopment takes place over a long period of time, as complete maturation and synaptogenesis are not finished until young adulthood. Understanding processes and mechanism of neurodevelopment is a prerequisite to understand the cellular and molecular basis of neurodevelopmental diseases such as autism or schizophrenia as well as toxicity-induced pathologies.

13.3 Main Principle of *In Vitro* Test Systems to Model DNT

Common methods to assess developmental neurotoxicity rely on high-dose testing in experimental animals. Although these test methods are labor and time intensive and have uncertainties in their prediction for human-specific toxicity, they are still required and accepted by regulatory authorities. During the past few decades, an immense improvement in the field of molecular biology, biotechnology, and bioinformatics has led to a better understanding of fundamental molecular and cellular mechanisms of human neurodevelopment. However, taking advantage of this knowledge to improve developmental neurotoxicity risk assessment for human health is a long-lasting process, which has started only a few years ago [90,91].

Turning away from the standard “black-box” animal experiment, the concept of pathways of toxicity (PoT) was introduced. According to this concept, and based on the assumption that cellular and molecular mechanisms and signaling pathways are shared by various biological processes, screening of chemicals is performed to identify interaction between substances of interest and pathways controlling important cellular functions. In the case of developmental neurotoxicity, the perturbation of specific pathways results in altered cellular functions

leading to impaired neurobehavioral and/or neuropathological effects. Alternatively, the concept of adverse outcome pathways (AOPs) is used. This concept is built on a link between a chemical of interest and a macromolecule resulting in an altered function, which is relevant for human risk assessment [92]. A prerequisite for these screening approaches is known pathways of toxicity or confirmed molecular targets. However, the mode of action of DNT-causing chemicals is mostly unknown and toxicity in cellular model systems can also be caused by nonselective interactions with biomolecules. To address this issue, the concept of toxicity endophenotype (TEP) has been developed [30,56]. In contrast to the above-described concepts, TEP uses a "backward approach." In the case of the nervous system, neurodevelopment is assumed to have resulted from a set of key developmental processes that can be recapitulated *in vitro*. Thus, to assess chemical-induced phenotypic changes, researchers have established *in vitro* assays to model such processes. They comprise proliferation and apoptosis, differentiation, migration, neuritogenesis, synaptogenesis, and myelination. Also desirable or undesirable functional changes in neuronal excitability or neuroinflammation can be modeled. Alterations in neurobehavioral functions due to exposure to xenobiotics can occur in the absence of neuropathological evidence of structural damage. This challenges the *in vitro* assessment, but recently the first attempts showed a successful translation of neurobehavioral endpoints of DNT into *in vitro* assays and readouts [31,93]. To replace the OECD test guideline TG426 for DNT testing, it would require an *in vitro* test battery of DNT responses to provide adequate data to predict the adverse effect of chemicals on human health [48,94]. In order to optimize an integrated set of tests as a basis for a reproductive/developmental test battery with a predictive power for toxicological safety assessment, ReProTect, an integrated project of the EU, was funded within the 6th Framework Programme. At the end of the project, a blinded testing approach showed that this novel approach for hazard and risk assessment of reproductive toxicity is feasible [95].

Using new testing approaches and technologies (e.g., transcriptomics or metabolomics) in a more integrated way, (developmental) neurotoxicity data will have a stronger scientific foundation and therefore can be extrapolated with a great certainty to expected human response [94,96]. This leads to an improved risk assessment for human health of industrial and environmental chemicals and hence supports regulatory decisions [97].

13.4 Requirements of an *In Vitro* Test System for DNT/NT

To replace the traditional animal-based tests, alternative *in vitro* test systems require a high predictivity of adverse neurotoxic effects in the (developing) brain. In comparison with animal tests, they should be faster, cheaper, easier to perform, and adaptable to high-throughput analysis (HTA) and high-content analysis (HCA). In addition, the sensitivity and specificity of the alternative *in vitro*

methods should be higher than those of the classical *in vivo* test according to the guideline TG426.

Since the past few years, human stem cells (hSCs) are a promising source of human cells for mechanistically oriented DNT/NT safety assessment [27,31,46,98]. There is growing evidence that cells from human origin improve the predictivity of adverse effects on the developing human brain [99,100]. Due to their extensive proliferation potential, hSCs offer the opportunity for large-scale production of human cells for HTA and HCA [29]. Over the years, an enormous gain of knowledge in the field of stem cell biology, developmental biology, and neurobiology has contributed to constantly growing numbers of SC-based *in vitro* test systems for (developmental) neurotoxicity assessment [101]. Much work has been done on comparison of *in vitro* mouse ESC differentiation with embryonic and fetal mouse *in vivo* development [102]. For example, global gene expression analyses have shown similarities between *in vitro* neuronal differentiation of embryonic stem cells and *in vivo* embryonic neural tube development. Moreover, the authors observed an apicobasal cell polarity, active Notch signaling, and accurate development of neurons and glia similar to the *in vivo* equivalent [103,104]. Due to a lack of accessibility of human early development, only few data are available, comparing *in vitro* neuronal differentiation of human embryonic stem cells (hESCs) with their *in vivo* counterpart. Recently, it was demonstrated that hESCs could be differentiated *in vitro* toward medium-sized spiny neurons by undergoing a three-step protocol composed of induction, regionalization, and terminal differentiation. By means of protein expression analysis, it was shown that these phases resemble *in vivo* neurodevelopment of the ventral telencephalon [105]. Considering the fact that stem cell differentiation *in vitro* recapitulates crucial neurodevelopmental events *in vivo*, they are an ideal tool to assess adverse effects of chemicals on the developing and mature nervous systems (see Table 13.1) [68,106–109].

In vitro systems should cover crucial processes, which are essential to set up the human nervous system, such as cell proliferation, apoptosis, differentiation, migration, neuritogenesis, synaptogenesis, or neuronal excitability (Figure 13.1). Based on epidemiological studies and neurological diseases, growing evidence suggests that these processes, once destroyed, lead to impaired nervous system functions. Therefore, modeling these processes with *in vitro*-based test systems should provide a good alternative for DNT/NT assessment.

For a successful application of human SC-derived cells for DNT/NT testing, the *in vitro* test systems have to be thoroughly assessed [30]. The golden rule "Know the system you are working with and understand what affects it" is of prime importance. This is well described in the principles of good cell culture practice (GCCP) published by the European Union Reference Laboratory of Alternative Methods to Animal Testing (ECVAM) [150] or in more general test system guidance [151].

Before biological systems can serve as *in vitro* test systems, they have to be intensively characterized. Establishing a hSC-based *in vitro* system requires the assessment of pluripotency of stem cells and characterization of the differentiation process toward a defined cell type. Such a quality control is achieved by a

Table 13.1 *In vitro* models for (developmental) neurotoxicity: *in vivo* processes and the corresponding *in vitro* models and endpoints.

In vivo process[a)]	*In vitro* system[b)]	Endpoint[c)]	Reference[d)]
Proliferation and cell death	2D; the mouse embryonic stem cell test (EST)	Separation of cell death and differentiation	[110]
	2D; ReNcell CX (immortalized human neural progenitor cell line)	Proliferation (BrdU incorporation); viability (propidium iodide exclusion)	[111]
	3D; hNPCs (human neural progenitor cells)	Viability (CellTiter-Blue assay); cell death (LDH assay); apoptosis (caspase 3/7 activity, TUNEL assay); proliferation (measuring spheres diameter, counting the number of cells/sphere)	[112]
	3D; hESCs (WA09 and abBG02)	Proliferation (BrdU incorporation, determine neurosphere area); differentiation (neurite formation by measuring neurite length)	[113]
Differentiation	3D; differentiating hESCs toward neural progenitors and neuronal precursors within neurospheres	Gene expression (q-PCR)	[114]
	2D; differentiating hESCs toward NEPs	Gene expression (microarray expression profiling, q-PCR); quantification of GFP expression in hESCs expressing GFP under the Hes5 promoter cell line (flow cytometry)	[28]
	2D; differentiating hESCs toward neuronal progenitors	Gene expression (q-PCR)	[115]
	2D; hESCs	Metabolomics	[116,117]
	2D; EST 3D; whole embryo culture	Toxicogenomics	[118]
	2D; neural differentiation of mouse ESCs	Use of biological categories (GO)	[119]
	2D; neural differentiation of mouse ESCs	Use of defined intervention for rescue of effect and proof of pathway; demonstration of shift from one lineage to another	[120]
	2D; neural differentiation of mouse ESCs	Demonstration of shift from one lineage to another; link to disease-specific genes	[87]
	2D; differentiating hESCs toward NEPs	Highly resolved concentration response; use of superordinate biological processes for toxicity index	[121]
	2D; neural differentiation of mouse ESCs	Calcium signaling	[122]
Migration	2D; hESC-based neural crest cell	Counting cells in a beforehand area, based on Hoechst and calcein-AM staining	[81]

	3D; hNPCs	Measuring the distance of migration of neural progenitor cells out of neurospheres	[112]
Neuritogenesis	2D; teratocarcinoma-derived hNT2 cells	Neurite growth with automated microscopy	[123]
	2D; SH-SY5Y	Measuring neurite network formation	[124]
	2D; LUHMES cell line	Measuring simultaneously neurite growth and viability by automated high-content image analysis	[125,126]
	2D; primary cultures of embryonic rat sympathetic neurons derived from superior cervical ganglia (SCG)	Quantifying axonal and dendritic growth	[127,128]
Synaptogenesis	2D; *ex vivo* brain slices	Measuring electrical activity and post-synaptic protein level	[129]
	2D; *ex vivo* brain slices; PC12 cells 3D; hippocampal slice cultures	Measuring electrical activity and post-synaptic protein level; measuring vesicular catecholamine release and intracellular Ca^{2+}	[130,131]
	2D; dissociated cortical networks from embryonic rats	Measuring circuit assembly (MEA)	[132]
Synaptogenesis	2D; cocultures of developing cochlear explants and hESC-derived neural progenitors	Presynaptic contacts (ICC: synapsin 1)	[133]
	2D; rodent primary mixed cortical cultures	Morphometric measurements from automated image analysis (dendrite length, puncta per neuron, puncta per cell body, puncta per dendrite length)	[134]
Myelination	3D; rat reaggregating brain cell culture	Protein kinase C activation	[135]
	3D; rat reaggregating brain cell culture	Gene expression (q-PCR); cytotoxicity (LDH release); glucose consumption; the rate of total RNA synthesis	[136]
Neuronal excitability	2D; fetal rat cortical neuronal networks	Measuring electrical activity (MEA)	[137]
	2D; rat cortical cultures	Measuring spontaneous activity in networks of cortical neurons (multiwell MEA)	[138]
	2D; rat primary cortical neurons, rat hippocampal slice cultures 3D; rat aggregating brain cell cultures	Measuring electrical activity (MEA)	[139–141]
	2D; primary cultures of rat cortical neurons	Measuring spontaneous network activity (MEA); individual action potential spikes, groups of spikes (bursts) in neuronal networks	[93]

(*continued*)

Table 13.1 (*Continued*)

In vivo process[a)]	*In vitro* system[b)]	Endpoint[c)]	Reference[d)]
	2D; primary cortical cultures	Analytical "fingerprinting" using PCA and chemical class prediction using support vector machines (SVMs) to classify chemical effects based on MEA data from 16 chemicals	[142]
	2D; hESC-derived neuronal cell networks	Measuring spontaneous electrical network activity (MEA)	[143]
Glia support of neurons[e)]	3D; rat aggregating brain cell culture	Gene expression (q-PCR); activity of microglia and astrocytes (ICC); MAP kinase pathway activation (immunoblotting)	[144]
	2D; mouse ESC-derived astrocytes	CD95 ligand-induced apoptosis, NO production, IL-6 expression, NF-κB activation, gene expression (q-PCR)	[145]
	2D; human monocyte-derived macrophages, rat microglial cells	Phagocytic recognition and removal of dying cells	[146]
	2D; mouse oligodendrocytes isolated from a primary mixed brain cell culture	Oligodendrocyte–microglia interaction in neuroinflammation	[147]
	2D; primary murine astrocytes; microglial cell line BV-2	Astrocyte–microglia interaction in neuroinflammation	[148,149]

a) Key biological processes of human nervous system development that are modeled *in vitro* and tested for disturbance by chemicals.
b) Examples for stem cell-based *in vitro* test systems based on the corresponding *in vivo* process. Also, a few test systems using rodent cells are listed to indicate what types of models may be used in the future based on human pluripotent cell-derived cultures.
c) Analytical endpoints are listed where relevant together with the type of disturbed process that is intended to be measured.
d) Exemplary work without intention to be complete or fully representative.
e) Models of neuroinflammation (failed or inverted glial support) and of phagocytic removal (overall tissue function of glia) are included here.

set of specific markers relying on different methods and readouts [98]. For instance, gene expression profiling is a common tool to carefully assess the biological system by analyzing cells at several time points during differentiation. This is crucial to determine which *in vivo* developmental stage and which fundamental biological process are modeled by the system. Based on these data, an appropriate set of functional markers and their adequate level of expression over time are defined. Furthermore, the characterization should include the determination of the time point of appearance and the level of enrichment of a desired progenitor cell population during *in vitro* differentiation [30,98,152]. If later

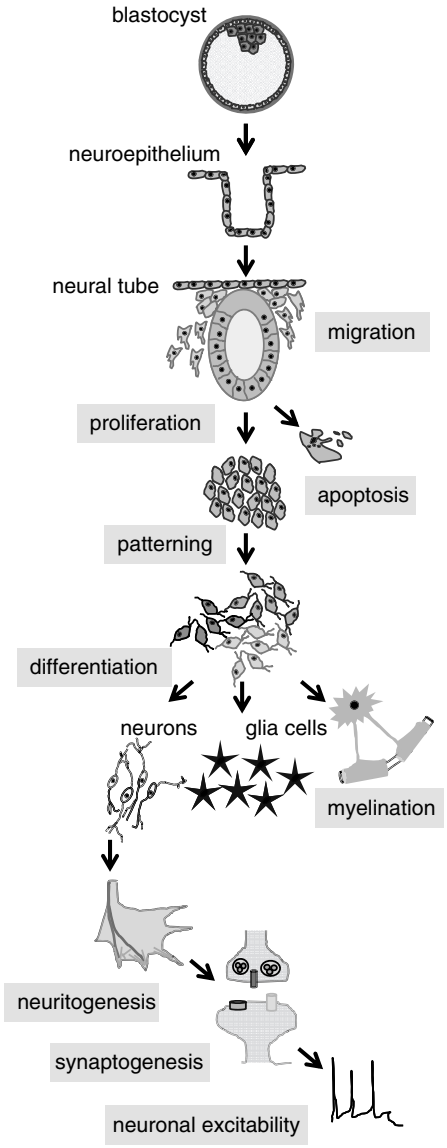


Figure 13.1 Important stages and processes during neurodevelopment. Schematic representation of crucial neurodevelopmental stages and processes *in vivo*, which can be modeled *in vitro* by differentiation of human pluripotent stem cells.

stages of neurodevelopment are supposed to be modeled, the functionality of the SC-derived neurons has to be proven by, for example, determination of electrophysiological activity. Further quality control measurements are quantification of axon formation, neurite growth, and presence of ion channels. They facilitate the creation of acceptability criteria for the use of individual cell batches for toxicity

testing and thereby increase the intra- and interlaboratory reproducibility [98]. The characterization of the system provides the basis for its usage in (developmental) neurotoxicity-specific endpoint assessment and helps define the right time window of exposure.

In the next step, the neurotoxicity-specific endpoints of the *in vitro* system have to be evaluated. One example is the "tandem approach," whereby the specificity of the assay is determined by comparing the response to a toxic compound and the response to a related nontoxic/less toxic chemical analog (substance tandem). Alternatively, the concept of "cell-type tandems" can be applied, which compares a sensitive cell type with a resistant cell type [30]. An alternative approach is to specifically block the toxicity of a compound by interference with an expected pathway of toxicity. This approach is still rarely used [120,153], although it produces a high level of confidence and plausibility. Since the quality of the new *in vitro* test system depends on a careful selection of tool compounds, the initial setup of an assay and the further testing is optimally performed with reference compounds that cover a range of chemical classes with known toxicity profiles and mechanisms. Two classes exist to determine the sensitivity of the test system: the "mechanistic tool compounds" (i.e., pathway-specific controls) and the "gold standard compounds" (i.e., compounds with proven effect on humans). The mechanistic tool compounds (e.g., pathway inhibitors) are known to alter the toxicity endpoint of interest. Therefore, they are also classified as "endpoint-specific controls." Gold standard compounds are highly important, as they are known to be (developmental) neurotoxic to humans, but their mode of action is often poorly characterized (e.g., methylmercury, arsenic, and lead compounds, ethanol, toluene, and polychlorinated biphenyls (PCBs)) [30,36]. To determine the specificity, negative control compounds are chosen that do not affect the endpoint that is quantified as readout of the test system [111,125,126,154]. Demonstrating that the test system is able to rapidly and efficiently screen a large number of chemicals with a high sensitivity and specificity, a larger set of test compounds should be chosen. For this, experts have made lists of chemicals that can be used for verification of the new test system. Based on several publications, and on data from humans and nonhuman primates as well as other laboratory mammals, chemicals are chosen that have an impact on nervous system function after developmental exposure [91]. In addition, lists are available for negative controls, pathway-specific compounds, and generally cytotoxic compounds [30]. To accelerate the development of new *in vitro* systems for DNT, experts from the field have published a document providing recommendations for developing alternative *in vitro* methods. In the publication, 15 guidance items have been outlined, which enable the evaluation and comparison of predictability and efficiency between several test assays and inter- and intralaboratory reproducibility [91]. A broader overview, also applying to neurotoxicity testing and other toxicological areas, is also available [151]. Due to the small number of compounds, also alternative strategies of compound collection have been explored, as, for instance, in the test battery designed by the EU ESNATS Consortium [155].

For regulatory acceptance, the *in vitro* test system still needs validation according to the guidelines, described by ICCVAM, OECD, and ECVAM [43,156–160]. The validation modules (test definition, within-laboratory variability (reliability), transferability (reliability), between-laboratory variability (reliability), predictive capacity (relevance), applicability (relevance), and performance standard) provide the required information to evaluate the validity of the test system [161,162], but new and faster forms of evaluation are being considered at the moment [48,163].

13.5 Modeling of Disease and Toxicant-Induced Damage

Conventionally, animal models are used to study mechanism and pathologies of human diseases. However, these model systems do not reflect the human *in vivo* situation. Many diseases do not occur in animals and have to be artificially induced [6]. For example, in animal models of toxicity-induced Parkinson's disease (PD), disease-specific symptoms and pathology are caused by injection of methylphenyltetrahydropyridine (MPTP), 6-hydroxydopamine (6-OHDA), paraquat, or rotenone [164]. Although *in vitro* neurotoxicology studies were performed to investigate the mode of action of environmental neurotoxins potentially playing a role in human neurodegeneration, the majority of them rely on models based on rodent primary cell cultures or have been performed in human (SH-SY5Y) or rodent (PC12) transformed cell lines that do not reflect the true state of neural cells *in vivo* [136,165]. Species-specific differences result in uncertainties in the extrapolation of results based on animal studies to the *in vivo* situation in humans. This may contribute to only a small number of therapeutic compounds that succeeded in clinical trials [166,167]. The breakthrough discovery of iPSCs has created new opportunities to study human development and disease directly in affected cell types [168,169]. Fibroblasts or other somatic cell types from healthy and diseased individuals can be converted into iPSCs, expanded, and differentiated into the desired neuronal cell type or alternatively directly converted into proliferating neural precursor cells (NPCs) or induced neuronal cells (iNCs) without an intervening stem or progenitor cell stage [170–172].

This circumvents species-specific differences and inaccessibility of neural tissue, and therefore provides a valuable tool to get new insights into human-specific disease mechanisms, progression, and pathology in different genetic backgrounds. In addition, it seems to be a promising platform for drug discovery and may enhance the likelihood for clinical success of new therapies [173–176].

A prerequisite for so-called "disease in a dish" models is an assessable alteration in the cellular or molecular phenotype in either the derived iPSCs or their differentiated progeny. Here, we discuss the current state of research with respect to neurological diseases. The most successful results have been achieved by studying monogenetic diseases, since a disease-associated genotype can be directly linked to the disease (cellular) phenotype. For example, spinal muscular

atrophy (SMA) is characterized by a progressive degeneration of motor neurons due to a mutation in the gene encoding for the survival of motor neuron protein. *In vitro* neuronal differentiation of SMA patient iPSC lines resulted in a specific loss of motor neurons, which could be rescued by VPA and tobramycin treatment similarly to the *in vivo* situation [177].

In another study, iPSCs were generated from Rett syndrome patients. *In vitro* differentiation of patient-specific iPSCs into GABA (γ-aminobutyric acid)-ergic neurons is characterized by disease-specific pathogenesis such as a decrease in the number of dendritic spikes and synapses, impaired intracellular calcium signaling, and electrophysiological defects [175,178,179].

Familiar dysautonomia (FD) is a disorder of the autonomic nervous system, which is caused by impaired neural crest cell migration due to a mutation in the gene encoding for IκB kinase complex-associated protein (IKBKAP). Generation of iPSCs from FD patients and *in vitro* differentiation into neural crest cells resulted in reduced migration and disrupted neurogenesis. Furthermore, it was shown that this model could be used as a discovery and screening platform to achieve new insights into the molecular and cellular mechanisms of the disease and to detect promising drug candidates to ameliorate migration and neuronal differentiation [11].

In addition, using iPSC technology to model spinocerebellar ataxia (Machado–Joseph disease (MJD)) demonstrated that reprogramming of patient-specific somatic cells and subsequent differentiation into the neuronal linage enables the study of late-onset neurodegenerative diseases in a cell-specific manner. Koch *et al.* highlighted that the disease-specific phenotype was exclusively found in neurons and was not present in iPSCs, glia, and fibroblasts [180].

For modeling Parkinson's disease (PD) *in vitro*, several iPSC lines have been generated from PD patients with defects in the LRRK2 gene. IPSC-derived midbrain dopaminergic neurons recapitulated several aspects of PD pathology *in vitro*, such as a reduced number of neurites and neurite branches per neuron, less resistance to oxidative stress, and impaired autophagy functions [181,182]. In addition, it could be shown that gene correction or inhibition of extracellular signal-regulated kinase 1/2 (ERK1/2) phosphorylation ameliorated neurodegeneration, neurite outgrowth, and dysregulation of genes, which are involved in other neurodegenerative pathologies [7].

Taken together, all these disease-specific iPSC lines are contributing to a better understanding of molecular and cellular mechanisms of diseases and link disease-associated genotypes to a disease-specific phenotype. However, a large percentage of neurological diseases are caused by the interaction of extrinsic factors and mutations in several genes. Despite the fact that it is much more challenging to develop "disease in a dish" models for complex nervous system disorders, the first models already have been published [173,176]. For instance, in 2011 Brennand *et al.* established a model for schizophrenia, a complex genetic psychiatric disorder. They identified crucial key signaling pathways of the disease phenotype. The disturbed pathways could be rescued by treatment with an antipsychotic drug [183]. Also, in a recent publication by Lancaster *et al.*, iPSCs

generated from one patient with microcephaly due to a mutation in the CDK5RAP2 gene were used to generate minibrains. The striking results were that the differentiation of these minibrains *in vitro* resembled the disease phenotype. The data also help to generate the hypothesis that the microcephalic phenotype developed due to premature neuronal differentiation at the expense of proliferation of neuronal progenitor cells [108].

One has to keep in mind that differences between the generated iPSC lines affect the differentiation and proper recapitulation of the disease. The cell lines vary, for instance, in their gene expression profile, genetic instability, X-inactivation, differentiation potential, and their epigenetic profile [184,185].

Another issue to be considered is that *in vitro*-derived neurons are often in an immature fetal stage and may lack full functionality, which may be suitable to investigate early-onset diseases *in vitro*. Indeed, to model late-onset diseases such as PD, cells were artificially aged by treating them with an environmental stressor to uncover a neurodegenerative phenotype in disease-derived dopaminergic neurons [6,7,181]. Alternatively, researchers are focusing on developing more complex long-term differentiation systems. The group of Livesey developed a human SC-based system that recapitulates crucial steps of the human cortical development *in vivo*, such as the formation of cortical stem cells and progenitor cells, a long period of neurogenesis (up to 100 days), a late phase of gliogenesis, acquirement of electrophysiological activity, synaptogenesis, and network formation. Since this system enables the generation of several classes of electrophysiologically active cortical projection neurons that form a functional excitatory synaptic network, it offers the potential to study complex neuropsychiatric disorders such as epilepsy, autism, schizophrenia, or neurodegenerative diseases [186].

Another major challenge for successful disease modeling is disease-associated phenotypes that emerge from non-cell autonomous interactions of two different cell types. For example, although one of the first patient-specific iPSC lines has been generated from amyotrophic lateral sclerosis (ALS), motor neurons derived from these cells lacked an obvious disease-specific phenotype [9]. However, coculture experiments with disease-specific glial cells (mutated in the SOD1 (superoxide dismutase 1) gene) revealed motor neuron degeneration [21].

The aim of the SC-based models is to recapitulate *in vitro* human neurodevelopment and diseases approximately closely to the *in vivo* equivalent (Figure 13.2). This emphasizes the need of developing *in vitro* three-dimensional (3D) systems to capture complex *in vivo* tissue physiology. In particular, during nervous system development cell–cell and cell–environment interactions play an essential role. The basic mechanisms of cell fate decisions during neurodevelopment rely on integration of external signals from extracellular matrix (ECM) components and gradients of mechanical and chemical stimuli from neighboring and distant cells [83]. Studies have demonstrated that gene expression profiles and cellular functionality of cells growing in 3D cell systems are more similar to the gene expression levels *in vivo* than to their two-dimensional (2D) counterpart [187]. First approaches to generate human nervous tissue *in vitro* already have been

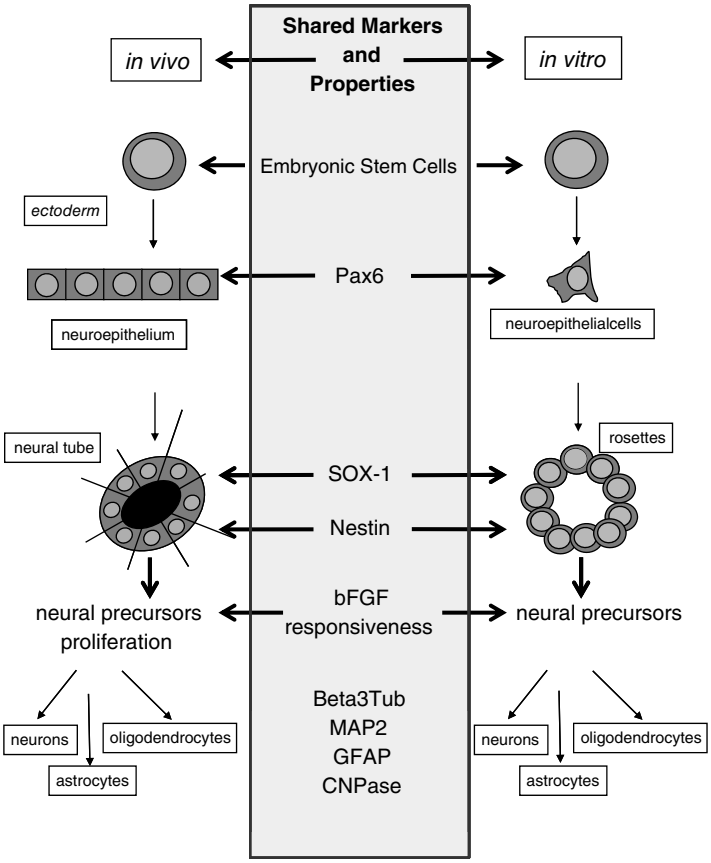


Figure 13.2 Differentiation of embryonic stem cells into neural cells *in vitro* mimics fetal brain development. The different neural stem cell populations that can be obtained *in vitro* correspond to stage-specific neural progenitors present at defined *in vivo* developmental stages and share the same markers and properties.

successfully performed, for example, by using an air–liquid interface system (Figure 13.3). Advantage of this technique is an improved gas exchange between tissue and air that enables differentiation into macroscopically 3D organized human neural-like tissue without hypoxic cell death [188]. *In vitro*-generated tissue consisted of neurons, astrocytes, and oligodendrocytes and resembled parts of early nervous system development [189]. Furthermore, a 3D SC-based system was established that recapitulates early dorsal telencephalic development in humans [190]. Recently, Lancaster *et al.* published the generation of cerebral organoids based on pluripotent SCs. This 3D organoid culture system enables the differentiation into different interdependent nervous system regions. Moreover, this system modeled microcephaly by differentiating patient-specific iPSCs and thereby contributed to a better understanding of the disease [108].

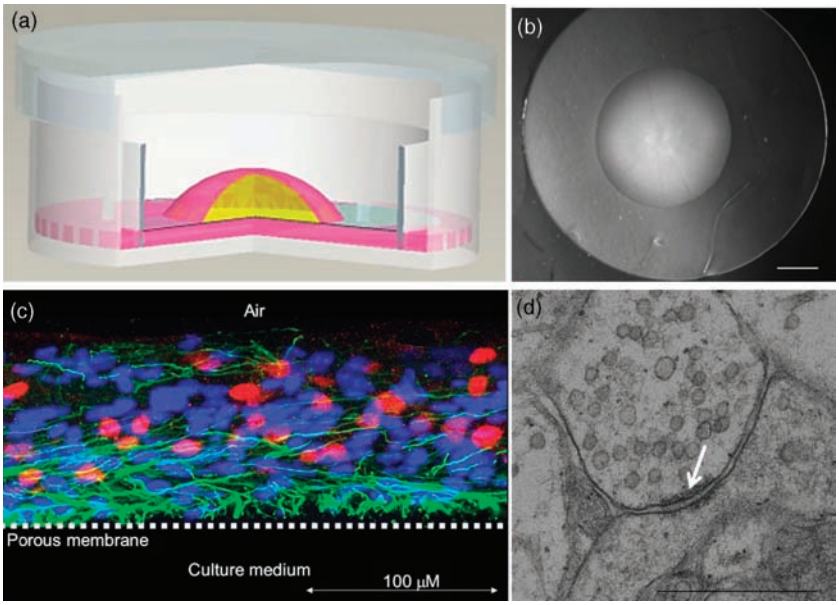


Figure 13.3 3D air/liquid neural cultures derived from human embryonic stem cells. (a) Scheme of 3D neural tissue grown onto porous membrane at the air/liquid interface. (b) Phase contrast photomicroscopy of a 6-month-old 3D neural tissue (bar = 500 mm). (c) Immunofluorescent staining of a transversal section of the 3D neural tissue shown in (b). (d) Electron microscopy microphotography of a 14-month-old 3D neural tissue showing a synapse (arrow).

Since the cellular milieu of the 3D tissue has been suggested to exhibit an *in vitro* microenvironment similar to the *in vivo* compartment, performance of toxicological and pharmacological screenings in these systems has the potential to identify *in vivo* toxicity. Forced by the need for quantitative and physiologically relevant *in vitro* systems, primary rat reaggregating brain cell cultures were used to assess neurotoxicity of heavy metals [191–194], neuron-specific toxicants [195–197], or organophosphorus compounds [198]. In addition, since these systems provide a microenvironment for key events during early neurodevelopment, they were also used to test for developmental neurotoxicity [192,199–202]. However, rat reaggregating brain cell cultures still bear the limitation of species differences. Currently, the differences are assessed between human and rodent SC-based systems [165]. Although human 3D ESC- or iPSC-based systems, or systems making use of committed fetal neural stem cells, enable the combination of *in vivo*-like complexity with an effective testing strategy, only a few systems have been published to assess relevant DNT endpoints such as viability, proliferation, migration, differentiation, neurite growth, and apoptosis [112,114,165,203,204]. However, due to the size and cell heterogeneity of 3D cell culture systems, it is more difficult to assess cell morphology and phenotype during differentiation, and therefore these systems are classified as low-throughput

systems [205]. Nevertheless, they provide a promising tool to investigate organ-specific developmental processes such as migration and differentiation, which should be incorporated into tiered testing strategies [165].

13.6 Using Stem Cells to Assess (Developmental) Neurotoxicity

13.6.1 Proliferation and Cell Death

During neurodevelopment, the process of neurogenesis and gliogenesis occurs highly coordinated directly after neural tube closure. The NEPs of the ventricular zone undergo symmetric cell division to increase the pool of NEPs. Later, they convert into radial glia (RG) and basal progenitors that are more restricted in their differentiation capacity. First, RG cells undergo symmetric cell division, followed by asymmetric cell division to produce neurons or glia [68,69]. Independent of their specific role during development, neural progenitor cells are responsible for the production of the correct number of a defined cell type (glia versus neurons; neuronal subtypes) in a spatially and temporally correct manner [206]. Besides proliferation, an enormous wave of cell death contributes to correct brain development. Thus, 20–80% (dependent on the population) of the developing cells undergo programmed cell death [207]. Since alteration of the cell number, due to intoxication, may result in nervous system abnormalities, proliferation and cell death are highly relevant biological endpoints [208,209]. Both can be modeled *in vitro*.

In one of the first attempts, an *in vitro* system based on the immortalized human neural progenitor cell line (ReNcell CX) was developed to assess chemical effects on proliferation (BrdU incorporation) and viability (propidium iodide exclusion) using high-throughput approaches. A training set of 16 compounds that included compounds known to cause DNT and noncytotoxic chemicals demonstrated the feasibility to screen chemicals for their DNT potential [111].

Besides 2D *in vitro* test systems, a few 3D human neurosphere systems have been developed that modeled key developmental processes such as proliferation, differentiation, migration, and apoptosis. These systems showed that neurotoxicant-induced modulations of these basic processes were quantifiable *in vitro*. They are able to assess the effect of potential DNT compounds on proliferation and apoptosis and provide the possibility to be used for medium-throughput screenings [112]. For example, an increase in proliferation potential has been measured in human neurospheres derived from ethanol-treated hESCs [113].

However, it has to be considered that compounds that have an effect on only a specific subgroup of cells cannot be detected by these traditional readouts. For example, neonatal treatment of rats with the antiproliferative agent methylazoxymethanol inhibited the proliferation of the neural progenitor cells but

simultaneously stimulated the survival of the newborn granular cells, resulting in an overall increase in cell numbers [210].

13.6.2 Differentiation

Although the detailed differentiation mechanism of individual cell fate specification may differ, they have some general principles in common. In particular, the tremendous diversity of neural cell types is based on the combination of spatially and temporally regulated lineage and environmental mechanisms. Once a cell has integrated all extrinsic and intrinsic cues at a defined time point during differentiation, specific transcriptional cascades are activated that regulate the corresponding program of differentiation [211].

Adverse effects of chemicals on neural differentiation are commonly analyzed by gene expression analysis [212]. Based on these data, alterations associated with a specific neurodevelopmental disease can be identified. It has been proven that this approach is useful to analyze time and concentration effects of chemicals, classify chemicals with regard to their toxicogenomic response, and compare *in vivo* and *in vitro* data as well as data among different species. Currently, guidelines are being developed for experimental and technical setup as well as for bioinformatics and statistical analysis to improve the quality of toxicogenomic studies. This will contribute to an integration of toxicogenomic approaches in the field of developmental neurotoxicity [212,213]. A smaller subgroup of thoroughly chosen marker genes has the potential to detect several developmental stages *in vitro* under physiological and toxic conditions [46,87,120]. Moreover, in contrast to normal cell lines, stem cell-based differentiation systems are dynamic, with continuously differentiating cells. This facilitates the assessment of chemicals that cause DNT by an "inhibition of a developmental function" in the absence of cytotoxicity [31]. Thus, changes in gene expression may be caused either by an acute effect of the chemical on a biological process, for example, alteration of signaling pathways, or by impaired differentiation.

To date, several *in vitro* systems have been established that recapitulate different stages of the early neural development. In most of the studies, methylmercury exposure or other manipulations (withdrawal of differentiation factors, addition of specific neural pathway inhibitors) were performed to demonstrate that disturbances of neural differentiation could be assessed by selected differentiation markers, such as neurite length, the neural transcription factor OTX2, or expression of neural genes such as TH or TUBB3 [28,114,115]. Using two separated exposure intervals enables the identification of the compound-specific sensitive phase of neurogenesis [115]. In addition, differentiating murine ESCs into morphologically and functionally mature neurons and glia-like cells offers the detection of cell type-specific toxicity by analyzing expression levels of cell type-specific marker genes. Even effects on small subpopulations and potential shifts of populations may be detected by this approach [120]. Recently, it was shown that the differentiation of hESCs toward NEPs was altered upon VPA treatment.

The observed transcriptional and epigenetic changes were linked to an altered neural development. Thus, the system might be used as a model to study drug-induced neurodevelopmental diseases [28]. Furthermore, an established 3D hESC-derived *in vitro* neurosphere system revealed distorted expression of neural differentiation markers due to polyethylene nanoparticle exposure. Therefore, the system may be used to investigate chronic nanotoxicity on developing NPCs and neuronal precursor cells, as the entire differentiation process takes over 32 days and can be followed based on multiple stage-specific markers [114]. Little attention has been paid to concentration-dependent effects yet. Several studies by the Piersma group used zebrafish or neurally differentiating murine ESCs for more than one concentration of a toxic compound, and some statistical analysis is available [119,214,215]. The first study approaching this phenomenon more comprehensively used human ESCs exposed to a large range of valproic acid concentrations. Interestingly, nontoxic, developmentally neurotoxic, and cytotoxic concentrations were clearly separated by gene expression patterns, and a DNT toxicity index was developed on this basis [121].

Proteomic as well as reporter gene technologies have been used to identify murine molecular toxicity signatures [216]. They may be transferred to a hESC-based system to screen for embryotoxicity and neurotoxicity in the future [217]. Furthermore, metabolomics has entered the field of toxicology. This technology enables the assessment of instantaneous alterations of cell homeostasis and provides information on toxicity-induced changes [96,218]. Studies in a hESC-based system showed that changes in levels of endogenous molecules due to teratogen exposures could predict human developmental toxicity [116,117]. Those studies indicate that this technology has the potential to identify DNT-specific biomarkers during neural differentiation of hESCs *in vitro* or primary reaggregating brain cell cultures.

During differentiation of hESCs toward neurons, the differentiating cells undergo several developmental stages with unique molecular and cellular characteristics that may contribute to differential sensitivity to certain compounds. Therefore, chemicals have to be tested during different developmental stages to facilitate the identification of the susceptible time window of specific DNT compounds [87].

13.6.3 Migration

Newly produced neurons and glia migrate from their place of origin to their final destination to establish a functional neuronal network and contribute to the architecture of the nervous system. Thus, specific neuronal populations have to undergo extensive radial and/or tangential migration over long distances to their distinct locations [219,220]. At the same time, also neural crest cells undergo extensive migration and follow stereotypical pathways to invade different types of tissues. All these processes are regulated by a multitude of extrinsic and intrinsic signals. To respond to guidance cues, the expression of appropriate

receptors at the right time is required [221]. For initiation, maintenance, and termination of migration, a coordinated assembly and rearrangement of actin and microtubule cytoskeleton is needed [222].

To evaluate potential DNT compounds that impair migration, an *in vitro* hSC-based test system was established to observe and quantify neural movement. For example, neurospheres consisting of fetal human neural stem cells were plated onto a coated cell culture surface to trigger migration of neural progenitor cells out of the neurosphere. The distance of migration was determined by measuring the distance between the edge of the neurosphere and the farthest migrated cell [4]. Environmental toxicants and inhibitors of physiological signaling pathways were also tested in a hESC-based neural crest cell system. Here migration of the cells was quantified by counting cells in a defined area. The results indicated a good sensitivity/specificity with regards to DNT compound identification [81]. Different alternatives to assess migration have been developed to enable the adaptation to high-throughput screenings.

13.6.4 Neuritogenesis

Correct and effective wiring of the nervous system requires the extension of axons and dendrites and finally their correct interconnection. After cells have migrated to their destination, they are exposed to a variety of signals from the extracellular environment that initiates neuritogenesis. In particular, gradients of guiding cues promote or inhibit neurite growth, branching, and adhesion. The underlining signals have to be transduced correctly into axonal growth patterns, mainly by cytoskeleton rearrangement. The growth cone, a specialized structure characterized by a dynamic cytoskeleton at the tip of axons, navigates the axons along a defined path to their appropriated target [223]. Alterations in this neurodevelopmental process have for a long time been suggested to be closely linked to DNT [36,224,225]. *In vitro* systems based on primary neurons and cell lines have been established using neurite growth as endpoint [123,125,126,154]. First approaches to assess neurite growth *in vitro* were based on low-density cultures. For example, teratocarcinoma-derived hNT2 cells were used to measure neurite growth with automated microscopy [123]. In another approach, SH-SY5Y neuroblastoma cells were plated onto a specialized surface to standardize the length of the neurite interconnections and neurite network formation was manually assessed [124]. More recent methods enabled the use of high-density cultures to measure neurite growth in LUHMES (Lund human mesencephalic) human neuronal precursor cells [226,227]. They have been used in an assay that simultaneously assessed neurite growth and viability by automated high-content image analysis. This approach allows the discrimination of compounds causing unspecific cytotoxicity and neurite-specific toxicity [125,126]. As automated microscopy is the underlying technique, the described test systems have the potential to be used in high-throughput screenings.

13.6.5 Synaptogenesis and Neuronal Excitability

As soon as the extending axon has reached an appropriated postsynaptic cell, the growth cone receives a defined signal to stop growing and to form a presynaptic terminal. At the same time, the target cell starts to build a postsynaptic site, which enables the establishment of a specialized connection between the cells, the synapse. Synapses transfer neuronal activity by transmitting patterns of electrochemical activity into neurotransmitter release [228]. One important feature of the human nervous system is the activity-dependent change in synaptic connectivity, called synaptic plasticity. It describes the mechanism of resculpting and rewiring of the neuronal network to alter thoughts, feelings, and behavior due to an experience [229]. The correct innervation of a given circuit requires that post- and presynaptic cells express the appropriate guidance receptors and intrinsic effectors to acquire the specific responsiveness to incoming signals [230]. The process of synaptogenesis includes the expression and the proper recruitment of specialized proteins for the stabilization of the initial synaptic contact, scaffold proteins for the specialized organization of both cells, rearrangement of the cytoskeleton, and endosomal tracking of synaptic growth signaling complexes. Disturbances in these processes result in early synapse loss, which is associated with several neuropsychiatric disorders such as autism, schizophrenia, and intellectual disabilities [3]. Moreover, neurons acquire the complete cellular and molecular equipment for functional activity. Excitability covers the process of signal reception, conversion, and transmission. Disturbances, for example, in the intracellular ion concentrations, in the integrity of ion channels and receptors, and in action potential generation, directly result in impaired neuronal excitability and can appear without other changes in biochemistry or morphology. Thus, neuronal excitability provides an important functional endpoint to assess (developmental) neurotoxicity [138]. Traditionally, synaptic plasticity has been analyzed in *ex vivo* brain slices by measuring electrical activity and postsynaptic protein levels [129,130]. Moreover, Ca^{2+} signaling is used as a functional endpoint *in vitro*, since intracellular Ca^{2+} concentrations regulate neurotransmission *in vivo* [231,232]. Recently, an *in vitro* model based on a coculture of hESC-derived neural progenitor cells with cochlear explants has been established to study synaptogenesis by measuring electrical activity and synaptic contact by immunohistochemistry [233]. Alternatively, microelectrode arrays (MEAs) are used to assess circuit assembly in different cell types [93,132,139,140]. As methods have been developed to assess synaptic plasticity *in vitro* that have the potential for higher throughput format, and hESC-based differentiation protocols emerge to model synaptogenesis and synaptic plasticity *in vitro*, a combination of both fields is expected in the future and will be beneficial for the establishment of new hSC-based *in vitro* models to detect DNT.

Furthermore, several studies have shown that MEA-based assays can be used to assess potential compounds affecting neuronal excitability and network formation with a high selectivity and sensitivity [137,138]. The MEA approach

enables the simultaneous recording of extracellular field potential in large populations of neuronal cells, which represents the spatial and temporal spike activities of multiple single neurons [133,234]. Electrophysiological activity is in fact one of the most sensitive and neuronal specific endpoints. It is providing high-information content on the neuronal tissue functional behavior. In the early 1980s, advancements in microfabrication technologies enabled the introduction of a new generation of devices, MEAs, which allowed *in vivo* [235] and *in vitro* [236,237] multisite, long-term recordings of the electrical activity of neuronal populations as well as the stimulation from one or more electrodes of the array. Planar substrates for *in vitro* experimentation, MEA-based neuroelectronic interfaces are now a well-accepted technique in basic and applied electrophysiology, enabling experimental investigations of collective dynamics, spatiotemporal patterns, and computational properties of neuronal assemblies in manners that were inaccessible before [238,239]. The use of this MEA approach for neurotoxicological studies [240] allows the distinction between neuronal cytotoxicity (irreversible damage and cell loss) and functional neurotoxicity where only the electrical signal propagation is affected by impaired or blocked transmission leaving cells metabolically intact. A recent study carried out by the group of Schafer on rat cortical neural networks shows that MEA data are useful for the separation of different chemical classes into effects classes (analytical "fingerprinting") of chemicals on neural networks [142].

The first series of neurotoxicological screening was performed using hESC-derived neuronal cell networks exhibiting spontaneous electrical activity by the group of Narkilahti [143]. Their results show that exposure for 72 h with 500 nM MeHgCl decreases the electrical signaling and alters the pharmacological response of hESC-derived neuronal networks in delayed manner, whereas effects cannot be detected with qRT-PCR, immunostaining, or proliferation measurements. The authors conclude that human cell-based MEA platform is a sensitive method for neurotoxicological screening. Human induced pluripotent stem cell (hiPSC)-derived neurons may also be effectively used as an effective method for drug discovery and neurotoxicological studies. However, compared with rat neurons, hiPSC-derived neurons seem to require longer time to mature functionally [241]. Finally, Suzuki *et al.* describe the development of planar carbon nanotube (CNT)-MEA chips that can measure both the release of the neurotransmitter dopamine and electrophysiological responses such as field post-synaptic potentials (fPSPs) and action potentials (APs). They successfully measured synaptic dopamine release from spontaneous firings with a high signal-to-noise (S/N) ratio. Therefore, these CNT-MEA chips might be accurate tools for *in vitro* assessment of potential neurotoxic effects of chemicals when used in combination with neural cell networks derived from human stem cells [242].

Moreover, measuring electrical activity in MEA-based assays has been shown to be a suitable endpoint for the detection of DNT-causing compounds after chronic exposure to low concentrations [139,140]. The production of multiwell MEAs and development of new methods for MEA data analysis facilitate

screening and prioritization of higher amounts of compounds for (developmental) neurotoxicity testing [243].

13.6.6 Myelination

Oligodendrocytes in the CNS and Schwann cells in the PNS generate myelin to wrap the axons of the neighboring neurons with a myelin sheath. This enables a fast and efficient flow of electrical impulses along myelinated axons and is essential for intact nervous system functions [244]. Interactions between axons and myelin-producing cells have several functions; for example, they control and modulate axonal growth, neuronal survival, and myelination [245,246]. As abnormalities in myelination are associated with several diseases such as multiple sclerosis, psychiatric diseases, and diverse peripheral neuropathies, several hESC-based *in vitro* protocols have been developed to generate oligodendrocytes and to model their interaction with neural and glial cells [112,189,205,247]. Studies in reaggregating brain cell cultures have shown that interactions between the different cell types (neurons, astrocytes, and oligodendrocytes) provide an environment facilitating neuronal differentiation leading to development of synapses and myelin [248]. Besides one study that solely investigates the effect of protein kinase C activation on myelination in 3D cultures [135], the cellular composition of the aggregates is usually assessed by analyzing changes in gene expression, as, for example, in the ACuteTox project [136]. Although human SC-based *in vitro* test systems exist that cover all three cell types of the brain, using myelination as an endpoint to assess (developmental) neurotoxicity in hESC-based *in vitro* systems is still challenging.

13.6.7 Neuroinflammation

Toxic compounds can directly or indirectly cause neurotoxicity. The term direct neurotoxicity describes phenomena in which the target cell of the toxic compound undergoes neurotoxicity-induced changes [249]. Any cell type of the brain can be the potential target of a certain compound. In contrast to direct toxicity, indirect neurotoxicity mainly occurs due to neuroinflammation caused by astrocytes and microglia [250]. Neuroinflammation is characterized by an increase in expression levels and/or release of proinflammatory cytokines (TNFα, IL-1β, and IL-6) and chemokines, reactivity of astrocytes and microglia, activation of inducible NO synthase (iNOS), and activation of MAPK and NF-κB signaling [251]. Several methods are available to assess these processes *in vitro*, for example, by analyzing the reactivity of microglia and astrocytes by immunocytochemistry, quantifying the expression levels of cytokines, chemokines, and iNOS on mRNA level, and determining the activation of MAP kinase pathway by Western blot [144,145,148,149]. In addition, gene expression of cytokines and chemokines as an endpoint for neuroinflammation would be adaptable to high-

throughput screening. However, since cell–cell interaction and maturational stages of the glial cells influence the neuroinflammatory response, an appropriate *in vitro* system has to be carefully chosen. To date, serum-free brain aggregate cell cultures are used to study neurotoxicity-induced neuroinflammation [250]. As protocols for human SC-based organoid cultures or neuron–glia cocultures are being established, and the feasibility to generate the cells in high quantities emerges, the methods to assess neurotoxicity-induced neuroinflammation can be easily transferred.

13.7 Limitations

Assessing (developmental) neurotoxicity in hSC-based *in vitro* test systems is contributing continually to a better understanding of molecular and cellular mechanisms of toxicity-induced phenotypic alterations. Since these *in vitro* test systems lack metabolic activity, only the primary toxic effect of a compound can be assessed. While *in vivo*, metabolism of xenobiotics by non-neural tissue (e.g., liver) or in the brain (e.g., astrocytes) may result either in a detoxification or in a bioactivation of the foreign compound. Thus, the *in vivo* metabolism of foreign compounds has to be carefully checked to avoid an underestimation or an overestimation of toxicity [32]. For example, *in vivo n*-hexane is biotransformed into the neurotoxic metabolite 2,5-hexanedione, parathion into paraxone, heroin deacetylated into morphine, and retinol into retinoic acid. Another example is MPTP, a model compound to induce neurotoxicity in animals. After administration, MPTP is converted by astrocytes into the neurotoxic compound MPP^+, which can be taken up by dopaminergic neurons [252]. Hence, an *in vitro* neurotoxicity test system may not be suitable for identification of parent compounds that act by active metabolites. However, there are different strategies to incorporate biotransformation [30].

In vitro test systems facilitate the detection of potential (developmental) neurotoxic compounds in a controlled system. However, compound actions in such test systems are isolated from physiological (*in vivo*) homeostasis mechanisms [32]. Since the prediction of neurotoxicity is related to the concentration at the target site, *in vitro* toxicokinetic models have to be developed. These also need to predict effects of the blood–brain barrier (BBB). For instance, the group of Cecchelli has developed a convenient approach to explore the importance of BBB permeability in neurotoxicity assessment of compounds by combining a BBB model with a neuronal cell line [222].

Besides the BBB, the placental barrier and binding to plasma proteins or to other non-neural tissue regulate *in vivo* chemical disposition. These defense mechanisms are mostly not present in *in vitro* test systems. Moreover, although hESC lines highly express the multidrug transporter ABCG2 to protect the cells against foreign toxic compounds, the differences in defense mechanisms influence the extrapolation of *in vitro* data to *in vivo* conditions. This has to be

included in the process of defining toxicologically relevant doses [32,253]. Predictions are also difficult, as the BBB is not fully developed until 6 months after birth. Moreover, the fetal liver does not exhibit the same detoxification mechanism as the adult liver [254].

From technical and practical points of view, physicochemical properties of compounds that influence the bioavailability have to be considered prior to *in vitro* testing, as it is challenging to evaluate chemicals that are insoluble in cell culture medium at neutral pH, or that are volatile and evaporate quickly. Moreover, direct chemical interaction may alter culture condition (pH shift, osmolarity), which influences the viability of the SC and their progeny. Binding of compounds to nutrients or other cell culture compounds may change the microenvironment of cells *in vitro* that do not mimic the *in vivo* situation [32].

Thus, the application domains for *in vitro* assays are still limited by technical constraints, and additional work is required to expand their utility for the prediction of a potential hazard toward the real prediction of risk [48].

Acknowledgments

This study was supported by grants and support from the Doerenkamp-Zbinden Foundation, the BMBF, the Land Baden-Württemberg, the Konstanz Research School Chemical Biology, the DFG (RTG1337), HES-SO, CTI/KTI, FP7 (ESNATS), and the SCAHT.

References

1 Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science*, **282** (5391), 1145–1147.

2 Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Harris, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.*, **227** (2), 271–278.

3 Ma, D.K., Bonaguidi, M.A., Ming, G.L., and Song, H. (2009) Adult neural stem cells in the mammalian central nervous system. *Cell Res.*, **19** (6), 672–682.

4 Moors, M., Cline, J.E., Abel, J., and Fritsche, E. (2007) ERK-dependent and -independent pathways trigger human neural progenitor cell migration. *Toxicol. Appl. Pharmacol.*, **221** (1), 57–67.

5 Colman, A. and Dreesen, O. (2009) Pluripotent stem cells and disease modeling. *Cell Stem Cell*, **5** (3), 244–247.

6 Merkle, F.T. and Eggan, K. (2013) Modeling human disease with pluripotent stem cells: from genome association to function. *Cell Stem Cell*, **12** (6), 656–668.

7 Reinhardt, P., Schmid, B., Burbulla, L.F., Schondorf, D.C., Wagner, L., Glatza, M., Hoing, S., Hargus, G., Heck, S.A., Dhingra, A., Wu, G., Muller, S., Brockmann, K., Kluba, T., Maisel, M., Kruger, R., Berg, D., Tsytsyura, Y., Thiel, C.S., Psathaki, O.E., Klingauf, J., Kuhlmann, T., Klewin, M., Muller, H., Gasser, T., Scholer, H.R., and Sterneckert, J. (2013) Genetic correction of a LRRK2 mutation in

human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell*, **12** (3), 354–367.

8 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126** (4), 663–676.

9 Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R., Wichterle, H., Henderson, C.E., and Eggan, K. (2008) Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science*, **321** (5893), 1218–1221.

10 Kaye, J.A. and Finkbeiner, S. (2013) Modeling Huntington's disease with induced pluripotent stem cells. *Mol. Cell. Neurosci.*, **56**, 50–64.

11 Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C.A., Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., Tabar, V., Sadelain, M., and Studer, L. (2009) Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. *Nature*, **461** (7262), 402–406.

12 Carvajal-Vergara, X., Sevilla, A., D'Souza, S.L., Ang, Y.S., Schaniel, C., Lee, D.F., Yang, L., Kaplan, A.D., Adler, E.D., Rozov, R., Ge, Y., Cohen, N., Edelmann, L.J., Chang, B., Waghray, A., Su, J., Pardo, S., Lichtenbelt, K.D., Tartaglia, M., Gelb, B.D., and Lemischka, I.R. (2010) Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature*, **465** (7299), 808–812.

13 Chow, M.Z., Boheler, K.R., and Li, R.A. (2013) Human pluripotent stem cell-derived cardiomyocytes for heart regeneration, drug discovery and disease modeling: from the genetic, epigenetic, and tissue modeling perspectives. *Stem Cell Res. Ther.*, **4** (4), 97.

14 Kaufman, D.S., Hanson, E.T., Lewis, R.L., Auerbach, R., and Thomson, J.A. (2001) Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA*, **98** (19), 10716–10721.

15 Nakamura, S., Takayama, N., Hirata, S., Seo, H., Endo, H., Ochi, K., Fujita, K.I., Koike, T., Harimoto, K.I., Dohda, T., Watanabe, A., Okita, K., Takahashi, N., Sawaguchi, A., Yamanaka, S., Nakauchi, H., Nishimura, S., and Eto, K. (2014) Expandable megakaryocyte cell lines enable clinically applicable generation of platelets from human induced pluripotent stem cells. *Cell Stem Cell*. doi: 10.1016/j.stem.2014.01.011.

16 Assady, S., Maor, G., Amit, M., Itskovitz-Eldor, J., Skorecki, K.L., and Tzukerman, M. (2001) Insulin production by human embryonic stem cells. *Diabetes*, **50** (8), 1691–1697.

17 Lumelsky, N., Blondel, O., Laeng, P., Velasco, I., Ravin, R., and McKay, R. (2001) Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science*, **292** (5520), 1389–1394.

18 Segev, H., Fishman, B., Ziskind, A., Shulman, M., and Itskovitz-Eldor, J. (2004) Differentiation of human embryonic stem cells into insulin-producing clusters. *Stem Cells*, **22** (3), 265–274.

19 Zhang, D., Jiang, W., Liu, M., Sui, X., Yin, X., Chen, S., Shi, Y., and Deng, H. (2009) Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. *Cell Res.*, **19** (4), 429–438.

20 Boissart, C., Poulet, A., Georges, P., Darville, H., Julita, E., Delorme, R., Bourgeron, T., Peschanski, M., and Benchoua, A. (2013) Differentiation from human pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric disease modeling and high-throughput drug screening. *Transl. Psychiatry*, **3**, e294.

21 Di Giorgio, F.P., Boulting, G.L., Bobrowicz, S., and Eggan, K.C. (2008) Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. *Cell Stem Cell*, **3** (6), 637–648.

22 Perrier, A.L., Tabar, V., Barberi, T., Rubio, M.E., Bruses, J., Topf, N., Harrison, N.L., and Studer, L. (2004)

Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA*, **101** (34), 12543–12548.

23 Bianco, P., Cao, X., Frenette, P.S., Mao, J.J., Robey, P.G., Simmons, P.J., and Wang, C.Y. (2013) The meaning, the sense and the significance: translating the science of mesenchymal stem cells into medicine. *Nat. Med.*, **19** (1), 35–42.

24 Daly, W., Yao, L., Zeugolis, D., Windebank, A., and Pandit, A. (2012) A biomaterials approach to peripheral nerve regeneration: bridging the peripheral nerve gap and enhancing functional recovery. *J. R. Soc. Interface*, **9** (67), 202–221.

25 Staflin, K., Honeth, G., Kalliomaki, S., Kjellman, C., Edvardsen, K., and Lindvall, M. (2004) Neural progenitor cell lines inhibit rat tumor growth *in vivo*. *Cancer Res.*, **64** (15), 5347–5354.

26 Schwartz, S.D., Hubschman, J.P., Heilwell, G., Franco-Cardenas, V., Pan, C.K., Ostrick, R.M., Mickunas, E., Gay, R., Klimanskaya, I., and Lanza, R. (2012) Embryonic stem cell trials for macular degeneration: a preliminary report. *Lancet*, **379** (9817), 713–720.

27 Leist, M., Bremer, S., Brundin, P., Hescheler, J., Kirkeby, A., Krause, K.H., Poerzgen, P., Puceat, M., Schmidt, M., Schrattenholz, A., Zak, N.B., and Hentze, H. (2008) The biological and ethical basis of the use of human embryonic stem cells for *in vitro* test systems or cell therapy. *ALTEX*, **25** (3), 163–190.

28 Balmer, N.V., Weng, M.K., Zimmer, B., Ivanova, V.N., Chambers, S.M., Nikolaeva, E., Jagtap, S., Sachinidis, A., Hescheler, J., Waldmann, T., and Leist, M. (2012) Epigenetic changes and disturbed neural development in a human embryonic stem cell-based model relating to the fetal valproate syndrome. *Hum. Mol. Genet.*, **21** (18), 4104–4114.

29 Cezar, G.G. (2007) Can human embryonic stem cells contribute to the discovery of safer and more effective drugs? *Curr. Opin. Chem. Biol.*, **11** (4), 405–409.

30 Kadereit, S., Zimmer, B., Thriel, C.van., Hengstler, J.G., and Leist, M. (2012) Compound selection for *in vitro* modeling of developmental neurotoxicity. *Front. Biosci. (Landmark Ed.)*, **17**, 2442–2460.

31 van Thriel, C., Westerink, R.H., Beste, C., Bale, A.S., Lein, P.J., and Leist, M. (2012) Translating neurobehavioural endpoints of developmental neurotoxicity tests into *in vitro* assays and readouts. *Neurotoxicology*, **33** (4), 911–924.

32 Harry, G.J., Billingsley, M., Bruinink, A., Campbell, I.L., Classen, W., Dorman, D.C., Galli, C., Ray, D., Smith, R.A., and Tilson, H.A. (1998) *In vitro* techniques for the assessment of neurotoxicity. *Environ. Health Perspect.*, **106** (Suppl. 1), 131–158.

33 EFSA (2013) Scientific opinion on the identification of pesticides to be included in cumulative assessment groups on the basis of their toxicological profile. *EFSA J.*, **11** (7), 3293.

34 Boyle, C.A., Decoufle, P., and Yeargin-Allsopp, M. (1994) Prevalence and health impact of developmental disabilities in US children. *Pediatrics*, **93** (3), 399–403.

35 Harada, M. (1995) Minamata disease: methylmercury poisoning in Japan caused by environmental pollution. *Crit. Rev. Toxicol.*, **25** (1), 1–24.

36 Grandjean, P. and Landrigan, P.J. (2006) Developmental neurotoxicity of industrial chemicals. *Lancet*, **368** (9553), 2167–2178.

37 Boyle, C.A., Boulet, S., Schieve, L.A., Cohen, R.A., Blumberg, S.J., Yeargin-Allsopp, M., Visser, S., and Kogan, M.D. (2011) Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, **127** (6), 1034–1042.

38 Grandjean, P. and Heindel, J.J. (2008) *In utero* and early-life conditions and adult health and disease. *N. Engl. J. Med.*, **359** (14), 1523. Author reply: 1524.

39 Hass, U. (2006) The need for developmental neurotoxicity studies in risk assessment for developmental toxicity. *Reprod. Toxicol.*, **22** (2), 148–156.

40 Makris, S.L., Raffaele, K., Allen, S., Bowers, W.J., Hass, U., Alleva, E., Calamandrei, G., Sheets, L., Amcoff, P.,

Delrue, N., and Crofton, K.M. (2009) A retrospective performance assessment of the developmental neurotoxicity study in support of OECD test guideline 426. *Environ. Health Perspect.*, **117** (1), 17–25.

41 U.S. EPA (1999) A set of scientific issues being considered by the Environmental Protection Agency regarding a retrospective analysis of developmental neurotoxicity studies. FIFRA SAP Report No. 99-01B, FIFRA Scientific Advisory Panel Meeting, 1998, Sheraton Crystal Hotel, Arlington, VA.

42 Goldman, L.R. and Koduru, S. (2000) Chemicals in the environment and developmental toxicity to children: a public health and policy perspective. *Environ. Health Perspect.*, **108** (Suppl. 3), 443–448.

43 OECD (2007) Test Guideline 426. OECD Guideline for Testing of Chemicals: Developmental Neurotoxicity Study. p. 26.

44 Middaugh, L.D., Dow-Edwards, D., Li, A.A., Sandler, J.D., Seed, J., Sheets, L.P., Shuey, D.L., Slikker, W., Jr., Weisenburger, W.P., Wise, L.D., and Selwyn, M.R. (2003) Neurobehavioral assessment: a survey of use and value in safety assessment studies. *Toxicol. Sci.*, **76** (2), 250–261.

45 Bottini, A.A., Amcoff, P., and Hartung, T. (2007) Food for thought . . . on globalisation of alternative methods. *ALTEX*, **24** (4), 255–269.

46 Kuegler, P.B., Zimmer, B., Waldmann, T., Baudis, B., Ilmjarv, S., Hescheler, J., Gaughwin, P., Brundin, P., Mundy, W., Bal-Price, A.K., Schrattenholz, A., Krause, K.H., van Thriel, C., Rao, M.S., Kadereit, S., and Leist, M. (2010) Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing. *ALTEX*, **27** (1), 17–42.

47 Krause, K.H., Thriel, C.van., De Sousa, P.A., Leist, M., and Hengstler, J.G. (2013) Monocrotophos in Gandaman village: India school lunch deaths and need for improved toxicity testing. *Arch. Toxicol.*, **87** (10), 1877–1881.

48 Leist, M., Hasiwa, N., Daneshian, M., and Hartung, T. (2012) Validation and quality control of replacement alternatives – current status and future challenges. *Toxicol. Res.*, **1** (1), 8–22.

49 Clancy, B., Finlay, B.L., Darlington, R.B., and Anand, K.J. (2007) Extrapolating brain development from experimental species to humans. *Neurotoxicology*, **28** (5), 931–937.

50 Gottmann, E., Kramer, S., Pfahringer, B., and Helma, C. (2001) Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environ. Health Perspect.*, **109** (5), 509–514.

51 Schardein, J.L., Schwetz, B.A., and Kenel, M.F. (1985) Species sensitivities and prediction of teratogenic potential. *Environ. Health Perspect.*, **61**, 55–67.

52 Wadman, M. (2006) Earlier drug tests on people could be unsafe, critics warn. *Nat. Med.*, **12** (2), 153.

53 Mol, P.G., Arnardottir, A.H., Motola, D., Vrijlandt, P.J., Duijnhoven, R.G., Haaijer-Ruskamp, F.M., de Graeff, P.A., Denig, P., and Straus, S.M. (2013) Post-approval safety issues with innovative drugs: a European cohort study. *Drug Saf.*, **36** (11), 1105–1115.

54 NRC (2007) *Toxicity Testing in the 21st Century. A Vision and a Strategy*. The National Research Council Committee on Toxicity Testing and Assessment of Environmental Agents, The National Academies Press, p. 196.

55 Leist, M., Hartung, T., and Nicotera, P. (2008) The dawning of a new age of toxicology. *ALTEX*, **25** (2), 103–114.

56 Balmer, N.V. and Leist, M. (2014) Epigenetics and transcriptomics to detect adverse drug effects in model systems of human development. *Basic Clin. Pharmacol. Toxicol.* doi: 10.1111/bcpt.12203.

57 Bal-Price, A.K., Hogberg, H.T., Buzanska, L., Lenas, P., van Vliet, E., and Hartung, T. (2010) *In vitro* developmental neurotoxicity (DNT) testing: relevant models and endpoints. *Neurotoxicology*, **31** (5), 545–554.

58 Bremer, S. and Hartung, T. (2004) The use of embryonic stem cells for regulatory developmental toxicity testing *in vitro* – the current status of test development. *Curr. Pharm. Des.*, **10** (22), 2733–2747.

59 Leist, M. (2013) Use of transcriptome profiling in stem cell based test systems for reproductive toxicity. LINZ 2013 – EUSAAT 2013 – ESNATS Conference, ALTEX, Linz, p. 65.

60 Massague, J. and Chen, Y.G. (2000) Controlling TGF-beta signaling. *Genes Dev.*, **14** (6), 627–644.

61 von Bubnoff, A. and Cho, K.W. (2001) Intracellular BMP signaling regulation in vertebrates: pathway or network? *Dev. Biol.*, **239** (1), 1–14.

62 Weinstein, D.C. and Hemmati-Brivanlou, A. (1997) Neural induction in *Xenopus laevis*: evidence for the default model. *Curr. Opin. Neurobiol.*, **7** (1), 7–12.

63 Pera, E.M., Wessely, O., Li, S.Y., and De Robertis, E.M. (2001) Neural and head induction by insulin-like growth factor signals. *Dev. Cell*, **1** (5), 655–665.

64 Wilson, S.I., Rydstrom, A., Trimborn, T., Willert, K., Nusse, R., Jessell, T.M., and Edlund, T. (2001) The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature*, **411** (6835), 325–330.

65 O'Rahilly, R. and Muller, F. (2007) The development of the neural crest in the human. *J. Anat.*, **211** (3), 335–351.

66 Smith, J.L. and Schoenwolf, G.C. (1997) Neurulation: coming to closure. *Trends Neurosci.*, **20** (11), 510–517.

67 Brand, A.H. and Livesey, F.J. (2011) Neural stem cell biology in vertebrates and invertebrates: more alike than different? *Neuron*, **70** (4), 719–729.

68 Conti, L. and Cattaneo, E. (2010) Neural stem cell systems: physiological players or *in vitro* entities? *Nat. Rev. Neurosci.*, **11** (3), 176–187.

69 Gotz, M. and Huttner, W.B. (2005) The cell biology of neurogenesis. *Nat. Rev. Mol. Cell Biol.*, **6** (10), 777–788.

70 Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011) Development and evolution of the human neocortex. *Cell*, **146** (1), 18–36.

71 Greene, N.D., Stanier, P., and Copp, A.J. (2009) Genetics of human neural tube defects. *Hum. Mol. Genet.*, **18** (R2), R113–R129.

72 Botto, L.D., Moore, C.A., Khoury, M.J., and Erickson, J.D. (1999) Neural-tube defects. *N. Engl. J. Med.*, **341** (20), 1509–1519.

73 Wallingford, J.B., Niswander, L.A., Shaw, G.M., and Finnell, R.H. (2013) The continuing challenge of understanding, preventing, and treating neural tube defects. *Science*, **339** (6123), 1222002.

74 Ornoy, A., Zvi, N., Arnon, J., Wajnberg, R., Shechtman, S., and Diav-Citrin, O. (2008) The outcome of pregnancy following topiramate treatment: a study on 52 pregnancies. *Reprod. Toxicol.*, **25** (3), 388–389.

75 Le Douarin, N.M. and Dupin, E. (2003) Multipotentiality of the neural crest. *Curr. Opin. Genet. Dev.*, **13** (5), 529–536.

76 Sauka-Spengler, T. and Bronner-Fraser, M. (2008) Evolution of the neural crest viewed from a gene regulatory perspective. *Genesis*, **46** (11), 673–682.

77 Menendez, L., Kulik, M.J., Page, A.T., Park, S.S., Lauderdale, J.D., Cunningham, M.L., and Dalton, S. (2013) Directed differentiation of human pluripotent cells to neural crest stem cells. *Nat. Protoc.*, **8** (1), 203–212.

78 Trainor, P.A. (2010) Craniofacial birth defects: the role of neural crest cells in the etiology and pathogenesis of Treacher Collins syndrome and the potential for prevention. *Am. J. Med. Genet. A*, **152A** (12), 2984–2994.

79 Taraviras, S. and Pachnis, V. (1999) Development of the mammalian enteric nervous system. *Curr. Opin. Genet. Dev.*, **9** (3), 321–327.

80 Amiel, J. and Lyonnet, S. (2001) Hirschsprung disease, associated syndromes, and genetics: a review. *J. Med. Genet.*, **38** (11), 729–739.

81 Zimmer, B., Lee, G., Balmer, N.V., Meganathan, K., Sachinidis, A., Studer, L., and Leist, M. (2012) Evaluation of developmental toxicants and signaling pathways in a functional test based on the migration of human neural crest cells.

Environ. Health Perspect., **120** (8), 1116–1122.

82 Menegola, E., Broccia, M.L., Di Renzo, F., Prati, M., and Giavini, E. (2000) *In vitro* teratogenic potential of two antifungal triazoles: triadimefon and triadimenol. *In Vitro Cell Dev. Biol. Anim.*, **36** (2), 88–95.

83 Briscoe, J. and Novitch, B.G. (2008) Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **363** (1489), 57–70.

84 Jessell, T.M. (2000) Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.*, **1** (1), 20–29.

85 Briscoe, J. and Therond, P.P. (2013) The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.*, **14** (7), 416–429.

86 Hui, C.C. and Angers, S. (2011) Gli proteins in development and disease. *Annu. Rev. Cell Dev. Biol.*, **27**, 513–537.

87 Zimmer, B., Kuegler, P.B., Baudis, B., Genewsky, A., Tanavde, V., Koh, W., Tan, B., Waldmann, T., Kadereit, S., and Leist, M. (2011) Coordinated waves of gene expression during neuronal differentiation of embryonic stem cells as basis for novel approaches to developmental neurotoxicity testing. *Cell Death Differ.*, **18** (3), 383–395.

88 Kamiya, A., Kubo, K., Tomoda, T., Takaki, M., Youn, R., Ozeki, Y., Sawamura, N., Park, U., Kudo, C., Okawa, M., Ross, C.A., Hatten, M.E., Nakajima, K., and Sawa, A. (2005) A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat. Cell Biol.*, **7** (12), 1167–1178.

89 Oksenberg, N. and Ahituv, N. (2013) The role of AUTS2 in neurodevelopment and human evolution. *Trends Genet.*, **29** (10), 600–608.

90 Bal-Price, A.K., Coecke, S., Costa, L., Crofton, K.M., Fritsche, E., Goldberg, A., Grandjean, P., Lein, P.J., Li, A., Lucchini, R., Mundy, W.R., Padilla, S., Persico, A. M., Seiler, A.E., and Kreysa, J. (2012) Advancing the science of developmental neurotoxicity (DNT): testing for better safety evaluation. *ALTEX*, **29** (2), 202–215.

91 Crofton, K.M., Mundy, W.R., Lein, P.J., Bal-Price, A., Coecke, S., Seiler, A.E., Knaut, H., Buzanska, L., and Goldberg, A. (2011) Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. *ALTEX*, **28** (1), 9–15.

92 Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount, D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrrano, J.A., Tietge, J.E., and Villeneuve, D.L. (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.*, **29** (3), 730–741.

93 Robinette, B.L., Harrill, J.A., Mundy, W.R., and Shafer, T.J. (2011) *In vitro* assessment of developmental neurotoxicity: use of microelectrode arrays to measure functional changes in neuronal network ontogeny. *Front. Neuroeng.*, **4**, 1.

94 Leist, M., Ringwald, A., Kolde, R., Bremer, S., Thriel, C.van., Krause, K.H., Rahnenfuhrer, J., Sachinidis, A., Hescheler, J., and Hengstler, J.G. (2013) Test systems of developmental toxicity: state-of-the-art and future perspectives. *Arch. Toxicol.*, **87** (12), 2037–2042.

95 Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., Schafer, W.R., Seiler, A., Witters, H., and Schwarz, M. (2010) The ReProTect Feasibility Study, a novel comprehensive *in vitro* approach to detect reproductive toxicants. *Reprod. Toxicol.*, **30** (1), 200–218.

96 Ramirez, T., Daneshian, M., Kamp, H., Bois, F.Y., Clench, M.R., Coen, M., Donley, B., Fischer, S.M., Ekman, D.R., Fabian, E., Guillou, C., Heuer, J., Hogberg, H.T., Jungnickel, H., Keun, H. C., Krennrich, G., Krupp, E., Luch, A., Noor, F., Peter, E., Riefke, B., Seymour, M., Skinner, N., Smirnova, L., Verheij, E., Wagner, S., Hartung, T., van Ravenzwaay, B., and Leist, M. (2013) Metabolomics in

toxicology and preclinical research. *ALTEX*, **30** (2), 209–225.

97 Keller, D.A., Juberg, D.R., Catlin, N., Farland, W.H., Hess, F.G., Wolf, D.C., and Doerrer, N.G. (2012) Identification and characterization of adverse effects in 21st century toxicology. *Toxicol. Sci.*, **126** (2), 291–297.

98 Pistollato, F., Bremer-Hoffmann, S., Healy, L., Young, L., and Stacey, G. (2012) Standardization of pluripotent stem cell cultures for toxicity testing. *Expert Opin. Drug Metab. Toxicol.*, **8** (2), 239–257.

99 Chapin, R.E. and Stedman, D.B. (2009) Endless possibilities: stem cells and the vision for toxicology testing in the 21st century. *Toxicol. Sci.*, **112** (1), 17–22.

100 Gassmann, K., Abel, J., Bothe, H., Haarmann-Stemmann, T., Merk, H.F., Quasthoff, K.N., Rockel, T.D., Schreiber, T., and Fritsche, E. (2010) Species-specific differential AhR expression protects human neural progenitor cells against developmental neurotoxicity of PAHs. *Environ. Health Perspect.*, **118** (11), 1571–1577.

101 Kang, K.S. and Trosko, J.E. (2011) Stem cells in toxicology: fundamental biology and practical considerations. *Toxicol. Sci.*, **120** (Suppl. 1), S269–S289.

102 Dubois-Dauphin, M.L., Toni, N., Julien, S.D., Charvet, I., Sundstrom, L.E., and Stoppini, L. (2010) The long-term survival of *in vitro* engineered nervous tissue derived from the specific neural differentiation of mouse embryonic stem cells. *Biomaterials*, **31** (27), 7032–7042.

103 Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., and Bekman, E. (2009) Neural differentiation of embryonic stem cells *in vitro*: a road map to neurogenesis in the embryo. *PLoS One*, **4** (7), e6286.

104 Nat, R. and Dechant, G. (2011) Milestones of directed differentiation of mouse and human embryonic stem cells into telencephalic neurons based on neural development *in vivo*. *Stem Cells Dev.*, **20** (6), 947–958.

105 Carri, A.D., Onorati, M., Lelos, M.J., Castiglioni, V., Faedo, A., Menon, R., Camnasio, S., Vuono, R., Spaiardi, P., Talpo, F., Toselli, M., Martino, G., Barker, R.A., Dunnett, S.B., Biella, G., and Cattaneo, E. (2013) Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32^{+} medium-sized spiny neurons. *Development*, **140** (2), 301–312.

106 Elkabetz, Y., Panagiotakos, G., Shamy, G.Al., Socci, N.D., Tabar, V., and Studer, L. (2008) Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes Dev.*, **22** (2), 152–165.

107 Fasano, C.A., Chambers, S.M., Lee, G., Tomishima, M.J., and Studer, L. (2010) Efficient derivation of functional floor plate tissue from human embryonic stem cells. *Cell Stem Cell*, **6** (4), 336–347.

108 Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013) Cerebral organoids model human brain development and microcephaly. *Nature*, **501** (7467), 373–379.

109 Pellizzer, C., Bremer, S., and Hartung, T. (2005) Developmental toxicity testing from animal towards embryonic stem cells. *ALTEX*, **22** (2), 47–57.

110 van Dartel, D.A., Zeijen, N.J., de la Fonteyne, L.J., van Schooten, F.J., and Piersma, A.H. (2009) Disentangling cellular proliferation and differentiation in the embryonic stem cell test, and its impact on the experimental protocol. *Reprod. Toxicol.*, **28** (2), 254–261.

111 Breier, J.M., Radio, N.M., Mundy, W.R., and Shafer, T.J. (2008) Development of a high-throughput screening assay for chemical effects on proliferation and viability of immortalized human neural progenitor cells. *Toxicol. Sci.*, **105** (1), 119–133.

112 Moors, M., Rockel, T.D., Abel, J., Cline, J.E., Gassmann, K., Schreiber, T., Schuwald, J., Weinmann, N., and Fritsche, E. (2009) Human neurospheres as three-dimensional cellular systems for developmental neurotoxicity testing.

Environ. Health Perspect., **117** (7), 1131–1138.

113 Krishnamoorthy, M., Gerwe, B.A., Scharer, C.D., Sahasranaman, V., Eilertson, C.D., Nash, R.J., Usta, S.N., Kelly, S., Rose, M., Peraza, R., Arumugham, J., Stewart, B., Stice, S.L., and Nash, R.J. (2013) Ethanol alters proliferation and differentiation of normal and chromosomally abnormal human embryonic stem cell-derived neurospheres. *Birth. Defects Res. B Dev. Reprod. Toxicol.*, **98** (3), 283–295.

114 Hoelting, L., Scheinhardt, B., Bondarenko, O., Schildknecht, S., Kapitza, M., Tanavde, V., Tan, B., Lee, Q.Y., Mecking, S., Leist, M., and Kadereit, S. (2012) A 3-dimensional human embryonic stem cell (hESC)-derived model to detect developmental neurotoxicity of nanoparticles. *Arch. Toxicol.*, **87** (4), 721–733.

115 Stummann, T.C., Hareng, L., and Bremer, S. (2009) Hazard assessment of methylmercury toxicity to neuronal induction in embryogenesis using human embryonic stem cells. *Toxicology*, **257** (3), 117–126.

116 Kleinstreuer, N.C., Smith, A.M., West, P.R., Conard, K.R., Fontaine, B.R., Weir-Hauptman, A.M., Palmer, J.A., Knudsen, T.B., Dix, D.J., Donley, E.L., and Cezar, G.G. (2011) Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.*, **257** (1), 111–121.

117 West, P.R., Weir, A.M., Smith, A.M., Donley, E.L., and Cezar, G.G. (2010) Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.*, **247** (1), 18–27.

118 Robinson, J.F., Theunissen, P.T., van Dartel, D.A., Pennings, J.L., Faustman, E.M., and Piersma, A.H. (2011) Comparison of MeHg-induced toxicogenomic responses across *in vivo* and *in vitro* models used in developmental toxicology. *Reprod. Toxicol.*, **32** (2), 180–188.

119 Theunissen, P.T., Schulpen, S.H., van Dartel, D.A., Hermsen, S.A., van Schooten, F.J., and Piersma, A.H. (2010) An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. *Reprod. Toxicol.*, **29** (4), 383–392.

120 Zimmer, B., Schildknecht, S., Kuegler, P.B., Tanavde, V., Kadereit, S., and Leist, M. (2011) Sensitivity of dopaminergic neuron differentiation from stem cells to chronic low-dose methylmercury exposure. *Toxicol. Sci.*, **121** (2), 357–367.

121 Waldmann, T., Rempel, E., Balmer, N.V., Konig, A., Kolde, R., Gaspar, J.A., Henry, M., Hescheler, J., Sachinidis, A., Rahnenfuhrer, J., Hengstler, J.G., and Leist, M. (2014) Design principles of concentration-dependent transcriptome deviations in drug-exposed differentiating stem cells. *Chem. Res. Toxicol.* doi: 10.1021/tx400402j.

122 Visan, A., Hayess, K., Sittner, D., Pohl, E.E., Riebeling, C., Slawik, B., Gulich, K., Oelgeschlager, M., Luch, A., and Seiler, A.E. (2012) Neural differentiation of mouse embryonic stem cells as a tool to assess developmental neurotoxicity *in vitro*. *Neurotoxicology*, **33** (5), 1135–1146.

123 Harrill, J.A., Freudenrich, T.M., Machacek, D.W., Stice, S.L., and Mundy, W.R. (2011) Quantitative assessment of neurite outgrowth in human embryonic stem cell-derived hN2 cells using automated high-content image analysis. *Neurotoxicology*, **31** (3), 277–290.

124 Frimat, J.P., Sisnaiske, J., Subbiah, S., Menne, H., Godoy, P., Lampen, P., Leist, M., Franzke, J., Hengstler, J.G., Thriel, C.van., and West, J. (2010) The network formation assay: a spatially standardized neurite outgrowth analytical display for neurotoxicity screening. *Lab Chip*, **10** (6), 701–709.

125 Krug, A.K., Balmer, N.V., Matt, F., Schonenberger, F., Merhof, D., and Leist, M. (2013) Evaluation of a human neurite growth assay as specific screen for developmental neurotoxicants. *Arch. Toxicol.*, **87** (12), 2215–2231.

126 Stiegler, N.V., Krug, A.K., Matt, F., and Leist, M. (2011) Assessment of chemical-induced impairment of human neurite outgrowth by multiparametric live cell imaging in high-density cultures. *Toxicol. Sci.*, **121** (1), 73–87.

127 Howard, A.S., Bucelli, R., Jett, D.A., Bruun, D., Yang, D., and Lein, P.J. (2005) Chlorpyrifos exerts opposing effects on axonal and dendritic growth in primary neuronal cultures. *Toxicol. Appl. Pharmacol.*, **207** (2), 112–124.

128 Kim, W.Y., Gonsiorek, E.A., Barnhart, C., Davare, M.A., Engebose, A.J., Lauridsen, H., Bruun, D., Lesiak, A., Wayman, G., Bucelli, R., Higgins, D., and Lein, P.J. (2009) Statins decrease dendritic arborization in rat sympathetic neurons by blocking RhoA activation. *J. Neurochem.*, **108** (4), 1057–1071.

129 Hernandez, R.V., Navarro, M.M., Rodriguez, W.A., Martinez, J.L., Jr., and LeBaron, R.G. (2005) Differences in the magnitude of long-term potentiation produced by theta burst and high frequency stimulation protocols matched in stimulus number. *Brain Res. Brain Res. Protoc.*, **15** (1), 6–13.

130 Dingemans, M.M., Ramakers, G.M., Gardoni, F., van Kleef, R.G., Bergman, A., Di Luca, M., van den Berg, M., Westerink, R.H., and Vijverberg, H.P. (2007) Neonatal exposure to brominated flame retardant BDE-47 reduces long-term potentiation and postsynaptic protein levels in mouse hippocampus. *Environ. Health Perspect.*, **115** (6), 865–870.

131 Fukunaga, K., Stoppini, L., Miyamoto, E., and Muller, D. (1993) Long-term potentiation is associated with an increased activity of Ca^{2+}/calmodulin-dependent protein kinase II. *J. Biol. Chem.*, **268** (11), 7863–7867.

132 Chiappalone, M., Massobrio, P., and Martinoia, S. (2008) Network plasticity in cortical assemblies. *Eur. J. Neurosci.*, **28** (1), 221–237.

133 Spira, M.E. and Hai, A. (2013) Multi-electrode array technologies for neuroscience and cardiology. *Nat. Nanotechnol.*, **8** (2), 83–94.

134 Harrill, J.A., Robinette, B.L., and Mundy, W.R. (2011) Use of high content image analysis to detect chemical-induced changes in synaptogenesis *in vitro*. *Toxicol. In Vitro*, **25** (1), 368–387.

135 Pouly, S., Storch, M.K., Matthieu, J.M., Lassmann, H., Monnet-Tschudi, F., and Honegger, P. (1997) Demyelination induced by protein kinase C-activating tumor promoters in aggregating brain cell cultures. *J. Neurosci. Res.*, **49** (2), 121–132.

136 Zurich, M.G. and Monnet-Tschudi, F. (2009) Contribution of *in vitro* neurotoxicology studies to the elucidation of neurodegenerative processes. *Brain Res. Bull.*, **80** (4–5), 211–216.

137 Defranchi, E., Novellino, A., Whelan, M., Vogel, S., Ramirez, T., van Ravenzwaay, B., and Landsiedel, R. (2011) Feasibility assessment of micro-electrode chip assay as a method of detecting neurotoxicity *in vitro*. *Front. Neuroeng.*, **4**, 6.

138 McConnell, E.R., McClain, M.A., Ross, J., Lefew, W.R., and Shafer, T.J. (2012) Evaluation of multi-well microelectrode arrays for neurotoxicity screening using a chemical training set. *Neurotoxicology*, **33** (5), 1048–1057.

139 Hogberg, H.T., Sobanski, T., Novellino, A., Whelan, M., Weiss, D.G., and Bal-Price, A.K. (2011) Application of micro-electrode arrays (MEAs) as an emerging technology for developmental neurotoxicity: evaluation of domoic acid-induced effects in primary cultures of rat cortical neurons. *Neurotoxicology*, **32** (1), 158–168.

140 van Vliet, E., Stoppini, L., Balestrino, M., Eskes, C., Griesinger, C., Sobanski, T., Whelan, M., Hartung, T., and Coecke, S. (2007) Electrophysiological recording of re-aggregating brain cell cultures on multi-electrode arrays to detect acute neurotoxic effects. *Neurotoxicology*, **28** (6), 1136–1146.

141 Thiebaud, P., Rooij, N.F.de., Koudelka-Hep, M., and Stoppini, L. (1997) Microelectrode arrays for electrophysiological monitoring of hippocampal organotypic slice cultures.

IEEE Trans. Biomed. Eng., **44** (11), 1159–1163.

142 Mack, C.M., Lin, B.J., Turner, J.D., Johnstone, A.F., Burgoon, L.D., and Shafer, T.J. (2014) Burst and principal components analyses of MEA data for 16 chemicals describe at least three effects classes. *Neurotoxicology*, **40**, 75–85.

143 Yla-Outinen, L., Heikkila, J., Skottman, H., Suuronen, R., Aanismaa, R., and Narkilahti, S. (2010) Human cell-based micro electrode array platform for studying neurotoxicity. *Front. Neuroeng.* doi: 10.3389/fneng.2010.00111.

144 Monnet-Tschudi, F., Defaux, A., Braissant, O., Cagnon, L., and Zurich, M.G. (2011) Methods to assess neuroinflammation. *Curr. Protoc. Toxicol.*, **12**, 12.19.

145 Kuegler, P.B., Baumann, B.A., Zimmer, B., Keller, S., Marx, A., Kadereit, S., and Leist, M. (2012) GFAP-independent inflammatory competence and trophic functions of astrocytes generated from murine embryonic stem cells. *Glia*, **60** (2), 218–228.

146 Hirt, U.A., Gantner, F., and Leist, M. (2000) Phagocytosis of nonapoptotic cells dying by caspase-independent mechanisms. *J. Immunol.*, **164** (12), 6520–6529.

147 Simon, B.M., Malisan, F., Testi, R., Nicotera, P., and Leist, M. (2002) Disialoganglioside GD3 is released by microglia and induces oligodendrocyte apoptosis. *Cell Death Differ.*, **9** (7), 758–767.

148 Henn, A., Lund, S., Hedtjarn, M., Schrattenholz, A., Porzgen, P., and Leist, M. (2009) The suitability of BV2 cells as alternative model system for primary microglia cultures or for animal experiments examining brain inflammation. *ALTEX*, **26** (2), 83–94.

149 Henn, A., Kirner, S., and Leist, M. (2011) TLR2 hypersensitivity of astrocytes as functional consequence of previous inflammatory episodes. *J. Immunol.*, **186** (5), 3237–3247.

150 Coecke, S., Balls, M., Bowe, G., Davis, J., Gstraunthaler, G., Hartung, T., Hay, R., Merten, O.W., Price, A., Schechtman, L., Stacey, G., and Stokes, W. (Second ECVAM Task Force on Good Cell Culture Practice) (2005) Guidance on good cell culture practice. A report of the Second ECVAM Task Force on Good Cell Culture Practice. *Altern. Lab. Anim.*, **33** (3), 261–287.

151 Leist, M., Efremova, L., and Karreman, C. (2010) Food for thought . . . considerations and guidelines for basic test method descriptions in toxicology. *ALTEX*, **27** (4), 309–317.

152 De Kock, J. (2012) Focus on stem cells as sources of human target cells for *in vitro* research and testing. *ALTEX*, **1** (1), 541–548.

153 Schreiber, T., Gassmann, K., Gotz, C., Hubenthal, U., Moors, M., Krause, G., Merk, H.F., Nguyen, N.H., Scanlan, T.S., Abel, J., Rose, C.R., and Fritsche, E. (2010) Polybrominated diphenyl ethers induce developmental neurotoxicity in a human *in vitro* model: evidence for endocrine disruption. *Environ. Health Perspect.*, **118** (4), 572–578.

154 Radio, N.M. and Mundy, W.R. (2008) Developmental neurotoxicity testing *in vitro*: models for assessing chemical effects on neurite outgrowth. *Neurotoxicology*, **29** (3), 361–376.

155 Zimmer, B., Pallocca, G., Dreser, N., Foerster, S., Waldmann, T., Westerhout, J., Julien, S., Krause, K., van Thriel, C., Hengstler, J., Sachinidis, A., Bosgra, S., and Leist, M. (2014) Profiling of drugs and environmental chemicals for functional impairment of neural crest migration in a novel stem cell-based test battery. *Arch. Toxicol.*, **88** (5), 1109–1126.

156 Stokes, W., McFarland, R., Kulpa-Eddy, J., Gatewood, D., Levis, R., Halder, M., Pulle, G., Kojima, H., Casey, W., Gaydamaka, A., Miller, T., Brown, K., Lewis, C., Chapsal, J.M., Bruckner, L., Gairola, S., Kamphuis, E., Rupprecht, C.E., Wunderli, P., McElhinney, L., De Mattia, F., Gamoh, K., Hill, R., Reed, D., Doelling, V., Johnson, N., Allen, D., Rinckel, L., and Jones, B. (2012) Report on the international workshop on alternative methods for human and veterinary rabies

vaccine testing: state of the science and planning the way forward. *Biologicals*, **40** (5), 369–381.

157 ICCVAM (1997) Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM).

158 ICCVAM (2003) ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods.

159 OECD (2005) Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

160 Stokes, W.S. and Wind, M. (2010) Validation of innovative technologies and strategies for regulatory safety assessment methods: challenges and opportunities. *ALTEX*, **27** (3), 87–95.

161 Judson, R., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R.R., Whelan, M., Xia, M., Huang, R., Austin, C., Daston, G., Hartung, T., Fowle, J.R., 3rd, Wooge, W., and Tong, W., and Dix, D. (2013) Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. *ALTEX*, **30** (1), 51–56.

162 Hartung, T. (2007) Food for thought . . . on validation. *ALTEX*, **24** (2), 67–80.

163 Hartung, T., Hoffmann, S., and Stephens, M. (2013) Mechanistic validation. *ALTEX*, **30** (2), 119–130.

164 Blesa, J., Phani, S., Jackson-Lewis, V., and Przedborski, S. (2012) Classic and new animal models of Parkinson's disease. *J. Biomed. Biotechnol.*, **2012**, 845618.

165 Breier, J.M., Gassmann, K., Kayser, R., Stegeman, H., De Groot, D., Fritsche, E., and Shafer, T.J. (2010) Neural progenitor cells as models for high-throughput screens of developmental neurotoxicity: state of the science. *Neurotoxicol. Teratol.*, **32** (1), 4–15.

166 Mullard, A. (2012) Oncology trials gear up for high-throughput sequencing. *Nat. Rev. Drug Discov.*, **11** (5), 339–340.

167 Scannell, J.W., Blanckley, A., Boldon, H., and Warrington, B. (2012) Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.*, **11** (3), 191–200.

168 Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131** (5), 861–872.

169 Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318** (5858), 1917–1920.

170 Han, D.W., Tapia, N., Hermann, A., Hemmer, K., Hoing, S., Arauzo-Bravo, M.J., Zaehres, H., Wu, G., Frank, S., Moritz, S., Greber, B., Yang, J.H., Lee, H. T., Schwamborn, J.C., Storch, A., and Scholer, H.R. (2012) Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell*, **10** (4), 465–472.

171 Thier, M., Worsdorfer, P., Lakes, Y.B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nothen, M.M., Brustle, O., and Edenhofer, F. (2012) Direct conversion of fibroblasts into stably expandable neural stem cells. *Cell Stem Cell*, **10** (4), 473–479.

172 Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, **463** (7284), 1035–1041.

173 Bellin, M., Marchetto, M.C., Gage, F.H., and Mummery, C.L. (2012) Induced pluripotent stem cells: the new patient? *Nat. Rev. Mol. Cell Biol.*, **13** (11), 713–726.

174 Grskovic, M., Javaherian, A., Strulovici, B., and Daley, G.Q. (2011) Induced pluripotent stem cells – opportunities for disease modelling and drug discovery. *Nat. Rev. Drug Discov.*, **10** (12), 915–929.

175 Marchetto, M.C. and Gage, F.H. (2012) Modeling brain disease in a dish: really? *Cell Stem Cell*, **10** (6), 642–645.

176 Trounson, A., Shepard, K.A., and DeWitt, N.D. (2012) Human disease modeling

with induced pluripotent stem cells. *Curr. Opin. Genet. Dev.*, **22** (5), 509–516.

177 Ebert, A.D., Yu, J., Rose, F.F., Jr., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2009) Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature*, **457** (7227), 277–280.

178 Ananiev, G., Williams, E.C., Li, H., and Chang, Q. (2011) Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as *in vitro* disease model. *PLoS One*, **6** (9), e25255.

179 Farra, N., Zhang, W.B., Pasceri, P., Eubanks, J.H., Salter, M.W., and Ellis, J. (2012) Rett syndrome induced pluripotent stem cell-derived neurons reveal novel neurophysiological alterations. *Mol. Psychiatry*, **17** (12), 1261–1271.

180 Koch, P., Breuer, P., Peitz, M., Jungverdorben, J., Kesavan, J., Poppe, D., Doerr, J., Ladewig, J., Mertens, J., Tuting, T., Hoffmann, P., Klockgether, T., Evert, B.O., Wullner, U., and Brustle, O. (2011) Excitation-induced ataxin-3 aggregation in neurons from patients with Machado–Joseph disease. *Nature*, **480** (7378), 543–546.

181 Nguyen, H.N., Byers, B., Cord, B., Shcheglovitov, A., Byrne, J., Gujar, P., Kee, K., Schule, B., Dolmetsch, R.E., Langston, W., Palmer, T.D., and Pera, R.R. (2011) LRRK2 mutant iPSC-derived DA neurons demonstrate increased susceptibility to oxidative stress. *Cell Stem Cell*, **8** (3), 267–280.

182 Sanchez-Danes, A., Richaud-Patin, Y., Carballo-Carbajal, I., Jimenez-Delgado, S., Caig, C., Mora, S., Di Guglielmo, C., Ezquerra, M., Patel, B., Giralt, A., Canals, J.M., Memo, M., Alberch, J., Lopez-Barneo, J., Vila, M., Cuervo, A.M., Tolosa, E., Consiglio, A., and Raya, A. (2012) Disease-specific phenotypes in dopamine neurons from human iPS-based models of genetic and sporadic Parkinson's disease. *EMBO Mol. Med.*, **4** (5), 380–395.

183 Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., McCarthy, S., Sebat, J., and Gage, F.H. (2011) Modelling schizophrenia using human induced pluripotent stem cells. *Nature*, **473** (7346), 221–225.

184 Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., Gnirke, A., Eggan, K., and Meissner, A. (2011) Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell*, **144** (3), 439–452.

185 Hayden, E.C. (2011) Stem cells: the growing pains of pluripotency. *Nature*, **473** (7347), 272–274.

186 Shi, Y., Kirwan, P., Smith, J., Robinson, H.P., and Livesey, F.J. (2012) Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nat. Neurosci.*, **15** (3), 477–486.

187 Abbott, A. (2003) Neuroscience: genomics on the brain. *Nature*, **426** (6968), 757.

188 Sundstrom, L., Biggs, T., Laskowski, A., and Stoppini, L. (2012) OrganDots – an organotypic 3D tissue culture platform for drug development. *Expert Opin. Drug Discov.*, **7** (6), 525–534.

189 Preynat-Seauve, O., Suter, D.M., Tirefort, D., Turchi, L., Virolle, T., Chneiweiss, H., Foti, M., Lobrinus, J.A., Stoppini, L., Feki, A., Dubois-Dauphin, M., and Krause, K.H. (2009) Development of human nervous tissue upon differentiation of embryonic stem cells in three-dimensional culture. *Stem Cells*, **27** (3), 509–520.

190 Mariani, J., Simonini, M.V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A.M., Horvath, T.L., and Vaccarino, F.M. (2012) Modeling human cortical development *in vitro* using induced pluripotent stem cells. *Proc. Natl. Acad. Sci. USA*, **109** (31), 12770–12775.

191 Eskes, C., Juillerat-Jeanneret, L., Leuba, G., Honegger, P., and Monnet-Tschudi, F. (2003) Involvement of microglia–neuron interactions in the tumor necrosis factor-alpha release, microglial activation, and neurodegeneration induced by

trimethyltin. *J. Neurosci. Res.*, **71** (4), 583–590.

192 Monnet-Tschudi, F., Zurich, M.G., Riederer, B.M., and Honegger, P. (1995) Effects of trimethyltin (TMT) on glial and neuronal cells in aggregate cultures: dependence on the developmental stage. *Neurotoxicology*, **16** (1), 97–104.

193 Sorg, O., Schilter, B., Honegger, P., and Monnet-Tschudi, F. (1998) Increased vulnerability of neurones and glial cells to low concentrations of methylmercury in a prooxidant situation. *Acta Neuropathol.*, **96** (6), 621–627.

194 Zurich, M.G., Monnet-Tschudi, F., and Honegger, P. (1994) Long-term treatment of aggregating brain cell cultures with low concentrations of lead acetate. *Neurotoxicology*, **15** (3), 715–719.

195 Honegger, P., Pardo, B., and Monnet-Tschudi, F. (1998) Muscimol-induced death of GABAergic neurons in rat brain aggregating cell cultures. *Brain Res. Dev. Brain Res.*, **105** (2), 219–225.

196 Monnet-Tschudi, F., Hazekamp, A., Perret, N., Zurich, M.G., Mangin, P., Giroud, C., and Honegger, P. (2008) Delta-9-tetrahydrocannabinol accumulation, metabolism and cell-type-specific adverse effects in aggregating brain cell cultures. *Toxicol. Appl. Pharmacol.*, **228** (1), 8–16.

197 Pardo, B. and Honegger, P. (1999) High sensitivity of immature GABAergic neurons to blockers of voltage-gated calcium channels. *Brain Res. Dev. Brain Res.*, **115** (1), 9–16.

198 Zurich, M.G., Honegger, P., Schilter, B., Costa, L.G., and Monnet-Tschudi, F. (2004) Involvement of glial cells in the neurotoxicity of parathion and chlorpyrifos. *Toxicol. Appl. Pharmacol.*, **201** (2), 97–104.

199 Eskes, C., Honegger, P., Jones-Lepp, T., Varner, K., Matthieu, J.M., and Monnet-Tschudi, F. (1999) Neurotoxicity of dibutyltin in aggregating brain cell cultures. *Toxicol. In Vitro*, **13** (4–5), 555–560.

200 Monnet-Tschudi, F., Zurich, M.G., and Honegger, P. (1993) Evaluation of the toxicity of different metal compounds in the developing brain using aggregating cell cultures as a model. *Toxicol. In Vitro*, **7** (4), 335–339.

201 Zurich, M.G., Eskes, C., Honegger, P., Berode, M., and Monnet-Tschudi, F. (2002) Maturation-dependent neurotoxicity of lead acetate *in vitro*: implication of glial reactions. *J. Neurosci. Res.*, **70** (1), 108–116.

202 Zurich, M.G., Honegger, P., Schilter, B., Costa, L.G., and Monnet-Tschudi, F. (2000) Use of aggregating brain cell cultures to study developmental effects of organophosphorus insecticides. *Neurotoxicology*, **21** (4), 599–605.

203 Hill, E.J., Woehrling, E.K., Prince, M., and Coleman, M.D. (2008) Differentiating human NT2/D1 neurospheres as a versatile *in vitro* 3D model system for developmental neurotoxicity testing. *Toxicology*, **249** (2–3), 243–250.

204 Vangipuram, S.D., Grever, W.E., Parker, G.C., and Lyman, W.D. (2008) Ethanol increases fetal human neurosphere size and alters adhesion molecule gene expression. *Alcohol. Clin. Exp. Res.*, **32** (2), 339–347.

205 Martinez, Y., Dubois-Dauphin, M., and Krause, K.H. (2012) Generation and applications of human pluripotent stem cells induced into neural lineages and neural tissues. *Front. Physiol.*, **3**, 47.

206 Corbin, J.G., Gaiano, N., Juliano, S.L., Poluch, S., Stancik, E., and Haydar, T.F. (2008) Regulation of neural progenitor cell development in the nervous system. *J. Neurochem.*, **106** (6), 2272–2287.

207 Oppenheim, R.W. (1991) Cell death during development of the nervous system. *Annu. Rev. Neurosci.*, **14**, 453–501.

208 Adlard, B.P., Dobbing, J., and Sands, J. (1975) A comparison of the effects of cytosine arabinoside and adenine arabinoside on some aspects of brain growth and development in the rat. *Br. J. Pharmacol.*, **54** (1), 33–39.

209 Kuwagata, M., Takashima, H., and Nagao, T. (1998) A comparison of the *in vivo* and *in vitro* response of rat

embryos to 5-fluorouracil. *J. Vet. Med. Sci.*, **60** (1), 93–99.

210 Ciaroni, S., Cecchini, T., Ferri, P., Ambrogini, P., Cuppini, R., Lombardelli, G., Peruzzi, G., and Del Grande, P. (2002) Postnatal development of rat dentate gyrus: effects of methylazoxymethanol administration. *Mech. Ageing Dev.*, **123** (5), 499–509.

211 Vaccarino, F.M., Ganat, Y., Zhang, Y., and Zheng, W. (2001) Stem cells in neurodevelopment and plasticity. *Neuropsychopharmacology*, **25** (6), 805–815.

212 Krug, A.K., Kolde, R., Gaspar, J.A., Rempel, E., Balmer, N.V., Meganathan, K., Vojnits, K., Baquie, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R.M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S., Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J., Hothorn, L., Bremer, S., van Thriel, C., Krause, K.H., Hengstler, J.G., Rahnenfuhrer, J., Leist, M., and Sachinidis, A. (2013) Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Arch. Toxicol.*, **87** (1), 123–143.

213 Robinson, J.F., Pennings, J.L., and Piersma, A.H. (2012) A review of toxicogenomic approaches in developmental toxicology. *Methods Mol. Biol.*, **889**, 347–371.

214 Hermsen, S.A., Pronk, T.E., van den Brandhof, E.J., van der Ven, L.T., and Piersma, A.H. (2012) Concentration–response analysis of differential gene expression in the zebrafish embryotoxicity test following flusilazole exposure. *Toxicol. Sci.*, **127** (1), 303–312.

215 Hermsen, S.A., van den Brandhof, E.J., van der Ven, L.T., and Piersma, A.H. (2011) Relative embryotoxicity of two classes of chemicals in a modified zebrafish embryotoxicity test and comparison with their *in vivo* potencies. *Toxicol. In Vitro*, **25** (3), 745–753.

216 Kern, I., Xu, R., Julien, S., Suter, D.M., Preynat-Seauve, O., Baquie, M., Poncet, A., Combescure, C., Stoppini, L., Thriel, C.V., and Krause, K.H. (2013) Embryonic stem cell-based screen for small molecules: cluster analysis reveals four response patterns in developing neural cells. *Curr. Med. Chem.*, **20** (5), 710–723.

217 Klemm, M. and Schrattenholz, A. (2004) Neurotoxicity of active compounds – establishment of hESC-lines and proteomics technologies for human embryo- and neurotoxicity screening and biomarker identification. *ALTEX*, **21** (Suppl. 3), 41–48.

218 van Vliet, E., Morath, S., Eskes, C., Linge, J., Rappsilber, J., Honegger, P., Hartung, T., and Coecke, S. (2008) A novel *in vitro* metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine. *Neurotoxicology*, **29** (1), 1–12.

219 Bayer, S.A. and Altman, J. (1991) Development of the endopiriform nucleus and the claustrum in the rat brain. *Neuroscience*, **45** (2), 391–412.

220 Kriegstein, A. and Alvarez-Buylla, A. (2009) The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.*, **32**, 149–184.

221 Marin, O., Valiente, M., Ge, X., and Tsai, L.H. (2010) Guiding neuronal cell migrations. *Cold Spring Harb. Perspect. Biol.*, **2** (2), a001834.

222 Hallier-Vanuxeem, D., Prieto, P., Culot, M., Diallo, H., Landry, C., Tahti, H., and Cecchelli, R. (2009) New strategy for alerting central nervous system toxicity: integration of blood–brain barrier toxicity and permeability in neurotoxicity assessment. *Toxicol. In Vitro*, **23** (3), 447–453.

223 O'Donnell, M., Chance, R.K., and Bashaw, G.J. (2009) Axon growth and guidance: receptor regulation and signal transduction. *Annu. Rev. Neurosci.*, **32**, 383–412.

224 Alfano, D.P. and Petit, T.L. (1982) Neonatal lead exposure alters the dendritic development of hippocampal dentate granule cells. *Exp. Neurol.*, **75** (2), 275–288.

225 Choi, B.H., Kudo, M., and Lapham, L.W. (1981) A Golgi and electron-microscopic study of cerebellum in methylmercury-

poisoned neonatal mice. *Acta Neuropathol.*, **54** (3), 233–237.

226 Lotharius, J., Falsig, J., Beek, J.van., Payne, S., Dringen, R., Brundin, P., and Leist, M. (2005) Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. *J. Neurosci.*, **25** (27), 6329–6342.

227 Scholz, D., Poltl, D., Genewsky, A., Weng, M., Waldmann, T., Schildknecht, S., and Leist, M. (2011) Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. *J. Neurochem.*, **119** (5), 957–971.

228 Colon-Ramos, D.A. (2009) Synapse formation in developing neural circuits. *Curr. Top. Dev. Biol.*, **87**, 53–79.

229 Abbott, L.F. and Nelson, S.B. (2000) Synaptic plasticity: taming the beast. *Nat. Neurosci.*, **3** (Suppl.), 1178–1183.

230 Polleux, F., Ince-Dunn, G., and Ghosh, A. (2007) Transcriptional regulation of vertebrate axon guidance and synapse formation. *Nat. Rev. Neurosci.*, **8** (5), 331–340.

231 Dingemans, M.M., de Groot, A., van Kleef, R.G., Bergman, A., van den Berg, M., Vijverberg, H.P., and Westerink, R.H. (2008) Hydroxylation increases the neurotoxic potential of BDE-47 to affect exocytosis and calcium homeostasis in PC12 cells. *Environ. Health Perspect.*, **116** (5), 637–643.

232 Westerink, R.H. and Vijverberg, H.P. (2002) Toluene-induced, Ca^{2+})-dependent vesicular catecholamine release in rat PC12 cells. *Neurosci. Lett.*, **326** (2), 81–84.

233 Nayagam, B.A., Edge, A.S., Needham, K., Hyakumura, T., Leung, J., Nayagam, D.A., and Dottori, M. (2013) An *in vitro* model of developmental synaptogenesis using cocultures of human neural progenitors and cochlear explants. *Stem Cells Dev.*, **22** (6), 901–912.

234 Stoppini, L., Duport, S., and Correges, P. (1997) A new extracellular multirecording system for electrophysiological studies: application to hippocampal organotypic cultures. *J. Neurosci. Methods*, **72** (1), 23–33.

235 Wise, K.D., Najafi, K., and Drake, K.L. (1984) A multichannel probe for intracortical single-unit recording. *IEEE Trans. Biomed. Eng.*, **31**, 583–1583.

236 Gross, G.W., Rieske, E., Kreutzberg, G.W., and Meyer, A. (1977) A new fixed-array multi-microelectrode system designed for long-term monitoring of extracellular single unit neuronal activity *in vitro*. *Neurosci. Lett.*, **6** (2–3), 101–105.

237 Pine, J. (1980) Recording action potentials from cultured neurons with extracellular microcircuit electrodes. *J. Neurosci. Methods*, **2** (1), 19–31.

238 Rutten, W. (2001) 'Rules of conduct': some practical guidelines for testing motor-nerve conduction. *Arch. Physiol. Biochem.*, **109** (2), 191.

239 Rutten, W.L. (2002) Selective electrical interfaces with the nervous system. *Annu. Rev. Biomed. Eng.*, **4**, 407–452.

240 Novellino, A., Scelfo, B., Palosaari, T., Price, A., Sobanski, T., Shafer, T.J., Johnstone, A.F., Gross, G.W., Gramowski, A., Schroeder, O., Jugelt, K., Chiappalone, M., Benfenati, F., Martinoia, S., Tedesco, M.T., Defranchi, E., D'Angelo, P., and Whelan, M. (2011) Development of micro-electrode array based tests for neurotoxicity: assessment of interlaboratory reproducibility with neuroactive chemicals. *Front. Neuroeng.*, **4**, 4.

241 Odawara, A., Saitoh, Y., Alhebshi, A.H., Gotoh, M., and Suzuki, I. (2014) Long-term electrophysiological activity and pharmacological response of a human induced pluripotent stem cell-derived neuron and astrocyte co-culture. *Biochem. Biophys. Res. Commun.*, **443** (4), 1176–1181.

242 Suzuki, I., Fukuda, M., Shirakawa, K., Jiko, H., and Gotoh, M. (2013) Carbon nanotube multi-electrode array chips for noninvasive real-time measurement of dopamine, action potentials, and postsynaptic potentials. *Biosens. Bioelectron.*, **49**, 270–275.

243 Lefew, W.R., McConnell, E.R., Crooks, J.L., and Shafer, T.J. (2013) Evaluation of microelectrode array data using Bayesian

modeling as an approach to screening and prioritization for neurotoxicity testing. *Neurotoxicology*, **36**, 34–41.

244 Bunge, R.P. (1968) Glial cells and the central myelin sheath. *Physiol. Rev.*, **48** (1), 197–251.

245 Miller, R.H. (2002) Regulation of oligodendrocyte development in the vertebrate CNS. *Prog. Neurobiol.*, **67** (6), 451–467.

246 Nave, K.A. and Salzer, J.L. (2006) Axonal regulation of myelination by neuregulin 1. *Curr. Opin. Neurobiol.*, **16** (5), 492–500.

247 Erceg, S., Ronaghi, M., Oria, M., Rosello, M.G., Arago, M.A., Lopez, M.G., Radojevic, I., Moreno-Manzano, V., Rodriguez-Jimenez, F.J., Bhattacharya, S.S., Cordoba, J., and Stojkovic, M. (2010) Transplanted oligodendrocytes and motoneuron progenitors generated from human embryonic stem cells promote locomotor recovery after spinal cord transection. *Stem Cells*, **28** (9), 1541–1549.

248 Guentert-Lauber, B., Monnet-Tschudi, F., Omlin, F.X., Favrod, P., and Honegger, P. (1985) Serum-free aggregate cultures of rat CNS glial cells: biochemical, immunocytochemical and morphological characterization. *Dev. Neurosci.*, **7** (1), 33–44.

249 Chang, L.W. (1992) The concept of direct and indirect neurotoxicity and the concept of toxic metal/essential element interactions as a common biomechanism underlying metal toxicity, in *The Vulnerable Brain and Environmental Risks* (eds R.L. Isaacson and K.F. Jensen), Plenum Press, New York, p. 61–88.

250 Monnet-Tschudi, F., Zurich, M.G., and Honegger, P. (2007) Neurotoxicant-induced inflammatory response in three-dimensional brain cell cultures. *Hum. Exp. Toxicol.*, **26** (4), 339–346.

251 Frank-Cannon, T.C., Alto, L.T., McAlpine, F.E., and Tansey, M.G. (2009) Does neuroinflammation fan the flame in neurodegenerative diseases? *Mol. Neurodegener.*, **4**, 47.

252 Ransom, B.R., Kunis, D.M., Irwin, I., and Langston, J.W. (1987) Astrocytes convert the parkinsonism inducing neurotoxin, MPTP, to its active metabolite, MPP^+. *Neurosci. Lett.*, **75** (3), 323–328.

253 Apati, A., Orban, T.I., Varga, N., Nemeth, A., Schamberger, A., Krizsik, V., Erdelyi-Belle, B., Homolya, L., Varady, G., Padanyi, R., Karaszi, E., Kemna, E.W., Nemet, K., and Sarkadi, B. (2008) High level functional expression of the ABCG2 multidrug transporter in undifferentiated human embryonic stem cells. *Biochim. Biophys. Acta*, **1778** (12), 2700–2709.

254 Hyatt, M.A., Budge, H., and Symonds, M.E. (2008) Early developmental influences on hepatic organogenesis. *Organogenesis*, **4** (3), 170–175.

14
Stem Cell-Based Methods for Identifying Developmental Toxicity Potential

Jessica A. Palmer, Robert E. Burrier, Laura A. Egnash, and Elizabeth L.R. Donley

14.1
Introduction

There is an increasing need for reliable, high-throughput *in vitro* developmental toxicity screens in both the pharmaceutical and chemical industries. Establishing predictive human cell-based assays to aid in the early discovery-phase detection of potential developmental toxicants is strongly warranted as these tests could reduce product development time and costs. Human embryonic stem (hES) cells are a relevant *in vitro* model for developmental toxicity testing, since they are derived from the developing embryo. In addition, the pathways that are important for human development are active in hES cells. The combination of hES cell culture with metabolomics, which can measure the complete set of small molecules (metabolites) in a sample, creates a unique model system for measuring changes in metabolite levels following toxicant exposure. This approach not only provides a prediction of toxicity potential but also mechanisms of toxicity. This chapter will cover the current tests used to identify developmental toxicity potential, the need for new testing strategies, technologies that can be used to develop these strategies, and the alternatives in development that are based on hES cells.

14.2
Developmental Toxicity Screening: Past and Present

14.2.1
Definition and Scope of the Problem

There is evidence that people have been interested in teratology, or the study of structural birth defects, dating back to as early as 6500 BC, although the current field of teratology has origins in the 1930s, when the first reports of birth defects induced in a mammalian species were published [1,2]. Prior to these studies, it was believed that the mammalian embryo was protected from harmful

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

Table 14.1 Wilson's six principles of teratology.

1. Susceptibility to teratogenesis depends on the genotype of the conceptus and the manner in which this interacts with environmental factors
2. Susceptibility to teratogenic agents varies with the developmental stage at the time of exposure
3. Teratogenic agents act in specific ways (mechanisms) on developing cells and tissues to initiate abnormal embryogenesis (pathogenesis)
4. The final manifestations of abnormal development are death, malformation, growth retardation, and functional disorder
5. The access of adverse environmental influences to developing tissues depends on the nature of the influences (agent)
6. Manifestations of deviant development increase in degree as dosage increases from the no-effect to the totally lethal level

environmental factors by the placenta. The related field of developmental toxicology is broader, encompassing the study of the pharmacokinetics, mechanisms, pathogenesis, and adverse outcomes resulting from the exposure of the developing fetus to harmful agents or conditions. The term "developmental toxicology" was first presented in the 1970s by James Wilson, which was also when he proposed his "six principles of teratology" that are still accepted today (Table 14.1) [2–5]. Developmental toxicity is classified into four categories: (1) embryonic lethality, (2) dysmorphogenesis or structural abnormalities, (3) intrauterine growth restriction, and (4) functional toxicities. Some birth defects can be clinically diagnosed prior to or at birth, such as spina bifida, whereas others may not be identified until later in life (i.e., kidney malformations).

Approximately 3% of all infants in the United States are affected by a major defect at birth, increasing to 6–7% by 1 year of age [2,6]. Globally, it is estimated that every year 7.9 million children are born with a serious defect of genetic origin and hundreds of thousands more are born with birth defects caused by maternal exposure to environmental teratogens [7]. Major birth defects have been the leading cause of infant mortality in the United States for more than 20 years, leading to more than 20% of all infant deaths in 2011 [8,9]. Approximately 3% of all birth defects are attributed to *in utero* exposure to toxic chemicals and physical agents (including environmental factors) and 25% of birth defects are believed to be the result of a combination of genetic and environmental factors [5]. However, 50–65% of birth defects still have unknown causes [5,10].

It has been estimated that more than 70 000 man-made chemicals are circulating in the environment. This is in addition to the thousands of drugs on the market with hundreds more being added each year [11]. This extensive chemical space poses a significant risk to the safety of pregnant women and the developing fetus. Only a small percentage of these chemicals have been evaluated for developmental toxicity [5]. Approximately 4100 chemicals have been tested for teratogenicity potential in animals, of these 66% were nonteratogenic in the species

tested, 7% were teratogenic in two or more species, 18% were teratogenic in the majority of species tested, and 9% produced equivocal experimental results [12]. In contrast, only about 50 chemical or physical agents are known to cause developmental toxicity in humans [11,13]. The current developmental toxicity testing methods are not amenable to screening the backlog of chemicals due to cost and time, presenting a need for new high-throughput screens.

14.2.2 Historical Strategies and the Need for New Human-Based Models

The thalidomide tragedy in the 1960s increased public awareness of developmental toxicants and the importance of developmental toxicity testing. Prior to this, chemicals had been tested for toxicity in adult animals, and were only sporadically tested in pregnant animals. As a result of this disaster, more systematic requirements for developmental toxicity testing were put into place by regulatory agencies around the world. The current requirements for developmental toxicity testing have been in place for over 40 years and require a new test compound to be evaluated for effects on embryofetal development in two animal species, one rodent and one nonrodent species. Unfortunately, no animal species have been viewed as ideal for developmental toxicity testing and the majority of studies are conducted in rats and rabbits [14]. In 2008, nearly 12 million animals were used for experimental and other purposes in the Member States of the European Union [15]. Almost 9% of these animals were used for toxicology studies [15]. These preclinical models have varying degrees of concordance with observed developmental toxicity in humans, having approximately 70–80% concordance to known human teratogens [16]. These "segment II" *in vivo* animal models require a large number of animals, kilogram quantities of test compound, and are both time consuming and expensive. Although these animal models are, and have long been, considered the regulatory gold standard, differences in species response to a compound may lead to missed signals of developmental toxicity and biological misinterpretation.

There is a growing initiative to reduce and replace the use of animal models in toxicity testing, due to the ethical concerns regarding animal welfare and the cost of performing these tests. This is especially important in light of the current initiatives in Europe (Registration, Evaluation, and Authorization of Chemicals (REACH)) and the United States (Tox21) to screen thousands of chemicals currently in circulation for toxicity potential. In the report "Toxicity Testing in the 21st Century: A Vision and Strategy" (Tox21c), the United States National Research Council (NRC) presents a vision for the future wherein toxicity testing is done largely *in vitro* using human cell lines [17]. The opportunity for alternative toxicology models is favored by the REACH and Tox21 initiatives, which strongly support the use of *in vitro* testing for the first phase of compound analysis. Alternative toxicology models will not be able to replace experimental animals altogether because certain *in vivo* systemic interactions that are relevant for toxic outcomes cannot be anticipated *in vitro*. Nonetheless, *in vitro* testing

using human cell lines may serve as a better predictor for human response in certain cell types and contribute to refining and reducing the use of animals in research.

14.3 Pluripotent Stem Cells

14.3.1 Definition

Pluripotent stem (PS) cell research dates back almost 40 years to the early 1970s, when the first mouse embryonic carcinoma (EC) cell lines were established. These cells are derived from the stem cells of germline tumors (i.e., teratocarcinomas) and are able to differentiate into derivatives from each of the three germ layers: ectoderm, mesoderm, and endoderm [18,19]. The initial work in EC cells led to the isolation of embryonic stem (ES) cells from the inner cell mass (ICM) of mouse blastocysts in 1981 [20,21] and human blastocysts in 1998 [22]. ES cells are pluripotent, self-renewing cells that have the ability to recapitulate embryonic development *in vitro*, differentiating into numerous cell types derived from the three primary germ layers, as well as the extraembryonic tissues that support development (Figure 14.1) [19,23–25]. In addition, ES cell-derived somatic cells are able to function in a manner similar to their *in vivo* counterparts following transplantation into animal models [23,25,26]. This unique feature makes ES cells an invaluable resource for research pertaining to early human development and its critical pathways.

The groundbreaking development of induced pluripotent stem (iPS) cells from mouse somatic cells in 2006 [27] and human cells a year later [28,29] has opened the floodgates to new possibilities in pluripotent stem cell research and application, without the ethical controversies that surround hES cells. Since 2006, iPS cells have become one of the most competitive and fast-paced fields in life science research. Human iPS cells share many key properties with hES cells, such as morphology, pluripotency, self-renewal, and similar gene expression profiles [30]. The first human iPS cell lines were derived by reprogramming somatic cells with the retrovirus- or lentivirus-mediated induction of just four transcription factors: *Oct4*, *Sox2*, *Klf4*, and *c-Myc* [28] or *Oct4*, *Sox2*, *Nanog*, and *Lin28* [29]. Several new techniques to induce pluripotency in somatic cells have been developed with the goal of avoiding the integration of undesired viral DNA into the host genome. The new techniques include repeated transfection of plasmids or minicircle DNAs [31–33], the episomal approach [34], the use of excisable viral vector systems [35–37], as well as methods that use proteins [38–43], mRNA [44–47], microRNA [48,49], small molecules, or chemicals [50–59]. In comparison with mES and even hES cells, iPS cell research is still in its infancy

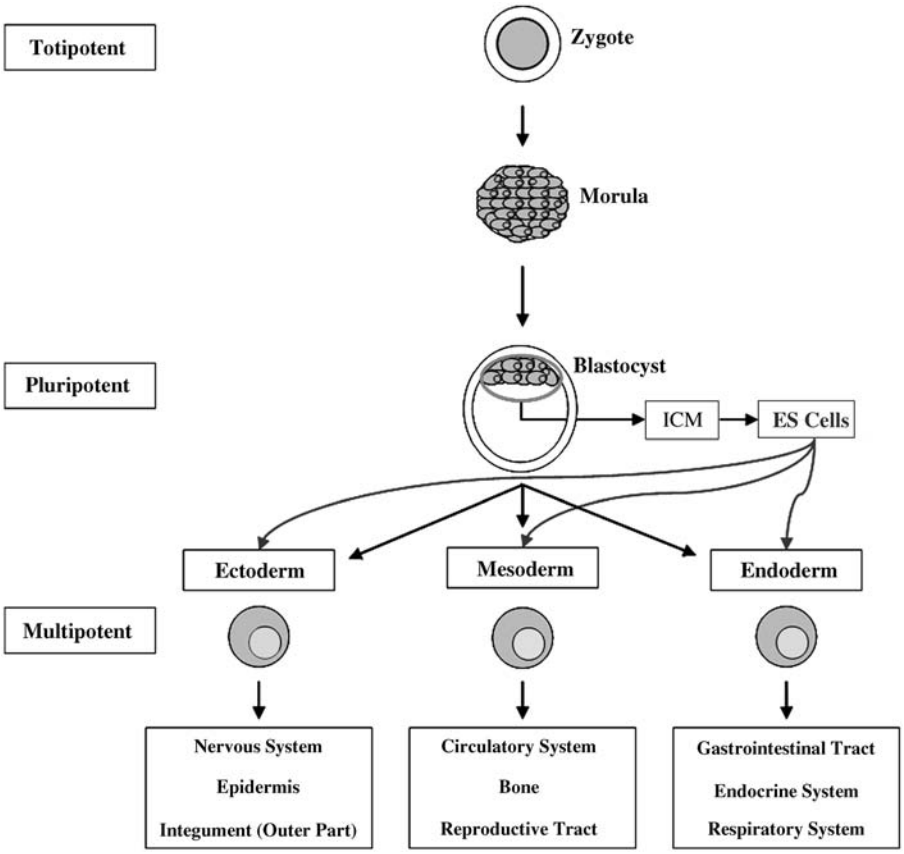


Figure 14.1 Embryonic stem cells are derived from the ICM of a blastocyst and differentiate into all cell types in the body.

and there is still a lot to learn. Recent research has revealed that numerous subtle, but important, molecular differences exist between ES and iPS cells [60]. These differences need to be fully understood before iPS cells can be viewed as a replacement for ES cells in research.

14.3.2 Ethical Considerations

The embryonic origin of human ES cells is the major reason that research in this field is a topic of great scientific interest and vigorous public debate, influenced by both ethical and legal positions. Derivation of hES cell lines requires the destruction of human embryos, which to some means destroying a potential human life [61]. Using hES cells has furthered our understanding of human development and disease in ways otherwise not possible.

14.4 Metabolomics

14.4.1 Definition

Metabolomics is the analysis of the complete set of small molecules (metabolites) present in a biological sample that are required for growth, maintenance, or normal function in a specific physiological state [62]. Metabolites are naturally occurring compounds with a molecular mass of less than 1 kDa that are diverse in their chemical structure and include lipids, sugars, and amino acids, to name a few [63]. Metabolites are generated from the action of enzymes and are the intermediates or end products of cellular regulatory processes. The metabolome is the entire collection of endogenous metabolites within a biological system and is analogous to the terms genome, transcriptome, and proteome. In addition, the metabolome can be measured at every level of complexity, that is, the organism as a whole, tissues, cells, or cell compartments [64]. Metabolites can be a part of the intracellular and/or extracellular metabolome (endometabolome and exometabolome, respectively) [65]. Analysis of the extracellular metabolites is also known as metabolic footprinting, whereas analysis of the intracellular metabolites is referred to as metabolic fingerprinting [66,67]. Gene expression and protein expression are predictors of what can happen in a cell, but the metabolites are a result of what is actually taking place, giving a more current view of a system's phenotype. The metabolite profile more closely reflects the present state of the cellular environment that is constantly changing due to nutrition, drug and pollutant exposures, and numerous other exogenous factors that influence health [63,68]. The concentration of metabolites is not simply a product of gene expression, but is the result of the interaction of the system's genome with its environment. The metabolic profile can provide important information about the health or disease state of the cellular sample [68]. The metabolome is more sensitive to perturbations than the transcriptome and proteome [67]. Metabolites can be mapped to biochemical pathways and serve as biomarkers for early identification of disease or chemical insult. They are biochemical features that can be altered and used to predict, diagnose, or measure the progress of disease or toxic response. In addition to biomarker identification, metabolomics can be used to identify the specific mechanisms of disease and toxicity. The measurement and analysis of the metabolic profile of a sample provides new insight into changes induced by external stimuli (i.e., drug treatment) and the metabolic phenotype [69].

14.4.2 Methods

Metabolite identification has been studied for more than 50 years, although the term "metabolomics" was not introduced until the late 1990s [68,70].

Metabolomic research has exploded over the past 10 years, due to the development of spectroscopic techniques with higher sensitivity, which is critical for metabolite identification. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy are most commonly employed analytical tools for metabolite profiling. NMR spectroscopy is advantageous for identifying structural information of metabolites, but is relatively insensitive when compared with MS (10^{-5} M versus 10^{-12} M) [69,71]. NMR sample analysis has a shorter acquisition time, but can typically identify only 20–50 metabolites per sample [63]. In addition, the amount of sample required for NMR is comparatively larger than that required for MS analysis (500 μl versus 10 μl, respectively) [63]. However, MS is not without its disadvantages. A major limitation of MS is the effects of ion suppression that arise from high concentrations of buffers and salts [63,72,73]. MS methods are also unable to differentiate chemical isomers with identical mass-to-charge ratios and provide little information on the structural components of the compounds identified. Tandem mass spectrometry (MS/MS) is able to provide more information on the structural characterization of metabolites, making it the "cornerstone" technique for metabolite identification [72]. The advantages and disadvantages of each method are summarized in Table 14.2.

Mass spectrometry is generally preceded by liquid chromatography (LC) or gas chromatography (GC), which separate the complex biological samples. In gas

Table 14.2 Comparison of techniques commonly used in metabolomic studies.

Detection technique	Separation technique	Advantages	Disadvantages
MS	Direct infusion	1. High throughput 2. Minimal sample preparation	1. Ion suppression limits detection
	Liquid chromatography	1. Broad coverage of metabolite classes 2. Gradient separations reduce ion suppression for ESI-MS 3. High sensitivity 4. Moderate cost	1. LC and MS parameters and instruments vary across laboratories 2. Can be biased by choice of sample preparation methods 3. Lengthy LC separations 4. Limited detection of volatiles
	Gas chromatography	1. Well developed 2. Standard databases for metabolite identification 3. High sensitivity 4. Moderate cost	1. Requires extensive sample preparation/derivatization
	Capillary electrophoresis	1. High chromatographic resolution 2. Sensitive	1. Underdeveloped technique 2. Lengthy CE separations
^{1}H NMR	Flow injection	1. Minimal sample preparation 2. Quantitative 3. Reproducible	1. Low sensitivity 2. Specialized staff required 3. Costly to purchase and operate

chromatography, samples must be derivatized prior to analysis. Derivatization chemically decreases the polarity of a compound to aid in separation on the column. Liquid chromatography is advantageous as it does not require this process [74,75]. The process of derivatization can be very harsh and adds the possibility of artifact formation, due to the alteration in compound structure. Use of LC–MS tools has increased in the field of metabolomics, as they provide higher sensitivity and a wider range for identifying the molecular mass of a compound [75,76]. In LC, the sample is introduced into a column, which can separate metabolites based on their chemical and physical interactions with the chromatographic packing material. The retention time of the metabolite (the time that it takes for a metabolite to elute) is a unique characteristic of each metabolite.

After separation on a LC column, a mass spectrometer identifies the mass and abundance of metabolites present in the sample. Atmospheric pressure ionization (API) sources have brought MS to the forefront of analytical tools for metabolomics, as they generate intact molecule ions at high sensitivity and allow coupling of MS instruments with normal- and reverse-phase chromatography [71,72]. There are a number of different API sources that can be used, such as electrospray ionization (ESI), atmosphere pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI). ESI-MS has become the method of choice for many researches for metabolite identification [73,77–79]. Prior to detection by the mass spectrometer, metabolites need to be converted into gas-phase ions. In ESI, this is accomplished by applying an electric field, which disintegrates a flowing liquid into a spray of fine, highly charged droplets [72]. The solvents quickly evaporate from these droplets before introduction into a mass analyzer [80]. This allows the mass-to-charge ratio (m/z) of solvent-free analyte ions to be determined. Ionization can be performed in positive or negative mode, which will produce protonated or deprotonated forms of the metabolite, respectively. The m/z that is measured can be converted into the molecular weight of the metabolite, aiding in metabolite identification.

14.4.3 Untargeted versus Targeted Metabolomic Approaches

Metabolomic methodologies fall into two distinct groups: targeted and untargeted. Targeted metabolomic experiments measure the concentrations of specific, chemically characterized and biochemically annotated metabolites, generally focusing on one or more related pathways of interest [81]. Developing hypotheses for these experiments requires researchers to utilize the comprehensive understanding of a vast array of metabolic enzymes, their kinetics, end products, and the known biochemical pathways to which they contribute [79]. Unlike untargeted metabolomics, targeted metabolomics can use isotopically labeled internal standards of the metabolite(s) of interest to quantitate a change in metabolite levels between biological samples. The use of internal standards for absolute quantitation helps to control the effects of ion suppression that can occur in untargeted metabolomic experiments [79]. The two most commonly

used approaches in targeted metabolomics are (1) selective extraction of the metabolites of interest in the biological sample and (2) setting up the detector system to selectively measure the metabolites of interest [81]. MS-based methods, followed by NMR, are the most commonly used approaches for targeted metabolomics because of their specificity and quantitative reproducibility [82]. Triple quadrupole (QqQ) MS methods are highly sensitive and robust, and are able to measure multiple metabolites with relatively high throughput. Additionally, they can quantitate low-concentration metabolites that are difficult to detect with less sensitive methods, such as NMR [82]. Although targeted metabolomic experiments are unable to identify unknown metabolites and metabolic perturbations, novel associations between the metabolites of interest may be illuminated in the context of specific physiological states.

The goal of untargeted metabolomics is to detect as many metabolites as possible within a specific range of mass values between different samples without bias [82,83]. Untargeted metabolomic experiments measure a wider breadth of metabolites, which can lead to the formation of new hypothesis based on metabolites that would not have been measured in a direct or targeted approach. Given the large number of metabolites measured, it is not possible to rely on isotopically labeled internal standards for quantitation, as is done in targeted metabolomic experiments. Quantitation is instead based on the mass ion intensity (i.e., the signal an ion generates at the detector) [83]. Since ion intensity can be impacted by other metabolites in a sample (i.e., ion suppression), it is important to compare samples that are as similar as possible. It is also important to try and analyze all samples on the same day using the same instrument and under the same set of conditions. This is particularly important in LC–MS experiments, which can have deviations in retention time between samples as a result of column degradation, sample carryover, and fluctuations in room temperature and mobile-phase pH [82]. Even with these drawbacks, LC–MS has become the method of choice in untargeted metabolomic experiments since it enables the detection of the most metabolites compared with other technologies (e.g., NMR) [82]. Using LC–MS, thousands of peaks, or metabolite features, can be identified in a biological sample, each having a unique m/z and retention time. Unfortunately, the structure and function of a large number of metabolites are still unknown and typically only about 25% of metabolite features can be tentatively identified [84]. These unknown metabolites can be identified using MS/MS and NMR methodologies; however, these experiments are time intensive. These drawbacks are outweighed by the potential of untargeted metabolomics to enhance our understanding of biochemistry and metabolism in biological systems.

14.4.4
Metabolomics in Toxicology

In the late 1990s, metabolomics was introduced as a new tool available to toxicologists. Since then, it has become an accepted methodology and incorporated as a part of the process aimed toward understanding toxicological profiles of

exogenous chemicals. This is expected to continue, particularly for compounds or classes of compounds where unexpected or unexplained findings arise, or if a specific finding calls for a deeper exploration of the biochemistry and pathophysiology. Since the metabolome closely reflects the activities of the cell at a functional level, metabolomic approaches can measure the final outcome after a toxic exposure in a cascade of events [85]. A metabolomic approach has the potential to rapidly identify a compound's potential targets as well as improve our understanding of the compound's mode of action [85]. Research groups in the pharmaceutical and chemical industries are using metabolomics in early *in vivo* toxicity screening to identify potential toxicological modes of action early in the compound discovery and development process, as reviewed by Robertson *et al.* [86] and Bouhifd *et al.* [87]. Other groups have utilized metabolomic approaches to understand the mechanisms of a compound-elicited toxic response and understand differences in response between genders, strains, age, and species [87]. An interesting study by Sumner *et al.* [88] measured the metabolite profiles in urine of pregnant CD rats and their pups following exposure to butylbenzyl phthalate (BBP). The metabolic profiles of the pups could differentiate male from female, pups exposed to the vehicle, low BBP dose, or high BBP dose, and pups with observable adverse reproductive effects from pups with no effects. The authors were able to map significant metabolites to biochemical pathways, providing insight into the mechanisms involved in BBP toxicity.

Metabolomic approaches continue to prove their utility and evolve from a basis for hypothesis generation to directed tools aimed at answering specific questions. The convergence of the metabolomic and toxicology fields has the potential to provide solutions toward meeting the sizeable challenge put forth by the NRC's Tox21c report [17]. It is envisioned that in the future the field of toxicology will come to rely more on human cell-based assays in the assessment of compound safety. To this end, metabolomics can help create a new generation of compound safety assays and continue to play a part as a methodology for understanding toxicity mechanisms. In striving to meet the goals of Tox21c, metabolomics can be used as a tool to identify model systems that can predict safety outcomes that in the future may be used together as a safety panel of assays.

Cell-based metabolomic approaches have already gained recognition in the toxicology field. These approaches are advantageous in that they are easier to control, less expensive, and easier to interpret than *in vivo* approaches [89]. In addition, both extracellular and intracellular metabolites can be measured, which can be associated with biomarkers from biofluids (extracellular samples) and provide information on the toxic mechanisms at a cellular level (intracellular) [87]. Intracellular metabolomic analyses have shown promising results in assessing neurotoxicity [90], renal toxicity [91], hepatotoxicity [92], mitochondrial toxicity [93], lung toxicity [94], and the toxicity of cell-penetrating peptides [95]. A major obstacle remains extrapolation of *in vitro* findings to *in vivo*. While the relationship between the two is still poorly understood, there are

ongoing efforts to better define this relationship and it is likely that a panel of cell-based assays will be needed to interpret results on a broader scale.

14.5
Stem Cell-Based *In Vitro* Screens for Developmental Toxicity Testing

14.5.1
Mouse Embryonic Stem Cell Test

The basis of using embryonic stem cells for developmental toxicity testing was pioneered in the 1990s with the development of the mouse embryonic stem cell test (mEST) [96]. This method combines changes in cardiomyocyte differentiation of mouse ES cells with differences in sensitivity between embryonic and adult tissues to cytotoxic damage in response to compound exposure. The mEST has been evaluated by the European Centre for Validation of Alternative Methods (ECVAM) and is considered a validated assay for embryotoxicity testing. However, the assay is relatively low throughput, requiring labor-intensive manipulations, long exposure periods, and the culture of two cell lines [97]. In addition, differentiation into specific lineages may limit an assay's potential for predicting teratogens that affect a different developmental lineage. Recent modifications to the mEST have been aimed at increasing throughput, adding additional developmental endpoints, and implementing molecular endpoints in place of subjective evaluation (reviewed in Ref. [98]). Multiple research groups have used a toxicogenomic approach (the application of genomics within toxicology) to understand the mechanisms of action of various developmental toxicants, such as triazoles [99] and valproic acid [100]. Additional studies have combined toxicogenomic methodologies and the mEST to build models for predicting developmental toxicity potential. These first initiatives focused on evaluating markers involved in cardiomyocyte differentiation [101,102]. Implementation of toxicogenomics-based assessments into the mEST using predetermined gene sets resulted in improved predictivity and the ability to discriminate between classes of developmental toxicants with distinct modes of action (reviewed in Refs [103,104]).

However, none of these modifications can overcome one of the biggest disadvantages of a developmental toxicity screen that utilizes animal cells, the possibility of species-specific differences in response to compound. There are multiple functional and morphological differences between hES and mES cells. To remain in an undifferentiated state, hES cells require basic fibroblast growth factor (bFGF), compared with leukemia inhibitory factor (LIF) in mES cells. In addition, hES cell cardiomyocyte differentiation takes longer and is less efficient than that for mES cells. In culture, hES cells grow as flat colonies, whereas mES cells form tight, rounded, multilayer colonies. The average population doubling time for hES cells is longer than that for mES cells (~30–35 h versus 12–15 h, respectively) [105]. Human ES cells form mainly cystic embryoid bodies (EBs), in

contrast to mES cells, which form both simple and cystic EBs. There are additional differences in the proteoglycans and stage-specific antigens expressed in hES and mES cells. These differences between hES and mES cell growth and differentiation indicate differences in key pathways during embryonic development, which could lead to differing response to compound exposure.

14.5.2 Human Embryonic Stem Cell-Based Developmental Toxicity Tests

The field of toxicology has been tasked by the NRC to transition from the traditional *in vivo* models that have been used for decades in toxicity testing to a new paradigm utilizing human cell-based assays [17]. Human ES cells and their derivatives hold great potential to provide functionally relevant cells for these new assays compared with traditional *in vitro* cell-based models, namely, immortalized cell lines and primary cells. Both immortalized cell lines and primary cells can be problematic in producing efficient and reproducible results. Immortalized cell lines do not represent normal cells found *in vivo*, while primary cells are expensive and have a very limited life span *in vitro* [106]. In comparison, hES cells are a more biologically relevant *in vitro* model system, especially for evaluating the developmental toxicity potential of chemicals. In addition, a developmental toxicity test based on hES cells reduces the risk of false negatives due specifically to interspecies differences in developmental pathways and pharmacokinetics [106,107]. The similarities between human ES and iPS cells may permit the use human iPS cells in place of hES cells in developmental toxicity assays, providing a less controversial model system. The use of hES cells in safety pharmacology and toxicology has received a lot of interest from both the pharmaceutical and chemical industries, which is illustrated by the collaborative projects and consortiums focused on developing hES cell-based assays, such as the consortiums Stem Cells for Safer Medicine (SC4SM; http://sc4sm.org/) and Embryonic Stem Cell-Based Novel Alternative Testing Strategies (ESNATS; www.esnats.eu/).

Several research groups have begun evaluating hES cells as a model for developmental toxicity testing. Proof-of-concept studies have been performed to evaluate the effects of various known developmental toxicants on the viability and integrity of undifferentiated hES cells and hES cell differentiation [108–118]. Adler *et al.* studied the effects of all-*trans*-retinoic acid, 13-*cis*-retinoic acid, valproic acid, dimethyl sulfoxide, and 5-fluorouracil on hES cells, hES cell-derived progenitors (excluding 5-fluorouracil), and human fibroblasts [108,109]. Since cytotoxicity alone is not sufficient for determining developmental toxicity potential, the authors developed a method for testing compound effects on hES cell cardiomyocyte differentiation by measuring gene expression changes. The gene markers identified in this study could be used as objective endpoints in a humanized version of the EST [108].

Another group developed an assay that assesses developmental toxicity potential by measuring cell proliferation following compound exposure (busulfan,

hydroxyurea, indomethacin, caffeine, penicillin G, and saccharin) in hES cells, hES cell-derived EBs, and human fibroblasts [111]. In addition, the assay evaluated the effects of compound on lineage-specific markers in differentiating EBs. A similar study by Pal *et al.* tested the effects of penicillin G, caffeine, and hydroxyurea on multiple endpoints, including hES cell adhesion and morphology as well as EB formation, viability, lineage-specific gene expression, and hormone secretion [117]. These studies demonstrate that hES cells can be used as a model to evaluate the embryotoxic effects of chemicals and that alterations in transcript levels of early lineage-specific markers can be correlated with known embryotoxic effects caused by drugs or chemicals. One of the drawbacks to these studies is the need for differentiation over an extended period of time. Unfortunately, differentiation can be variable in hES cells even when the same cell culture conditions are used. Moreover, these studies have evaluated only a very limited number of compounds, so while the assays show promise for developmental toxicity testing, it is still not known how predictive these methods will be in comparison with the currently employed *in vivo* and *in vitro* assays.

14.5.3
Combining Human Embryonic Stem Cells and Metabolomics: A Powerful Tool for Developmental Toxicity Testing

Metabolomics and hPS cell-based assays, independently, are exciting new approaches that can advance the field of developmental toxicity screening toward the directive put forth by the NRC's Tox21c report. Combining these two approaches offers a unique opportunity to understand the metabolic mechanisms involved in developmental toxicity at the earliest stages of human development [119–122], since hES cells are metabolically similar to embryonic epiblast cells at gastrulation. In addition, this combination can be used to build assays to predict the developmental toxicity potential of new chemical entities using a human-based model [121–123].

The developmental toxicity assays described here utilize an untargeted metabolomic platform that measures changes in hES cell metabolism in the spent cell culture medium (i.e., cell culture supernatant) following compound exposure, which is a functional measurement of cellular metabolism referred to as the secretome. The "secretome" is comprised of media components, metabolites passively and actively transported across the plasma membrane, and those produced through extracellular metabolism of enzymes. Changes in the secretome following compound exposure are a metabolic signature of toxicity that is related to alterations that occur both in the endometabolome (inside the cell) and in the extracellular matrix. A hPS cell-based assay combined with an untargeted metabolomic approach for developmental toxicity testing is advantageous for several reasons. (1) Changes in metabolite levels in response to toxicant exposure are a sensitive and quantitative measurement, enabling more objective data-driven decisions. (2) Multiple biochemical pathways can be assessed simultaneously, reinforcing the robustness of the

model when applied to drugs with a variety of mechanisms of toxicity. (3) Metabolic endpoints can be rapidly integrated with protein, DNA, and RNA targets for further pathway-based investigation. (4) Since the prediction is based on multiple independent variables, it is possible to detect teratogens exhibiting complex changes in metabolic patterns. (5) The assay measurements are predictive independent of cytotoxicity.

West *et al.* [122] performed the initial proof-of-concept study using low throughput, six-well tissue culture plates that are the standard culture platform for hES cells. The predictive model developed using these methods was 88% accurate in predicting the developmental toxicity potential of an independent test set of eight compounds [122]. However, the methods used in this study are not ideal for high-throughput evaluation of test compounds. The assay was migrated to 96-well plates to increase throughput and reproducibility. This new assay was evaluated using a blinded subset of 11 chemicals selected from the Environmental Protection Agency (EPA)'s ToxCast^{TM} chemical library. Using the predictive model trained on 23 pharmaceutical agents (i.e., the training set) of known human developmental toxicity, the assay predicted the blinded compounds with 73% accuracy in concordance with animal data. Retraining the model with data from one concentration level of each of the unblinded compounds increased the predictive accuracy to 83% for the remaining concentrations [121]. The untargeted metabolomics-based developmental toxicity assay is able to identify metabolic pathways that play a role in the test compound's mechanism of toxicity. A subset of the significantly enriched metabolic pathways is listed in Table 14.3. Many of these pathways had been previously suggested to underlie developmental toxicity. The results obtained in the untargeted metabolomics-based developmental toxicity assay can be used to identify a compound's or compound class' mechanisms of toxicity and aid in understanding the biochemical pathways affected by a toxic insult.

An untargeted metabolomic approach is able to simultaneously predict developmental toxicity with a high degree of accuracy and provide information on the metabolic pathways perturbed by the toxic compounds. However, it is still relatively low throughput given the amount of time required to analyze each sample using a lengthy chromatographic separation method and perform the data analysis and interpretation. To address this, a targeted biomarker-based assay was developed based on two predictive biomarkers, ornithine and cystine, that were identified in the untargeted metabolomic studies [123]. Ornithine is secreted by the hES cells, whereas cystine is present in the cell culture media. Changes in the biochemical pathways that contain ornithine and cystine as reactants or products have been experimentally associated with mechanisms of teratogenesis. Ornithine is formed as a product of the catabolism of arginine into urea, is critical to the excretion of nitrogen, and is a precursor to polyamines. Increased ornithine secretion inhibits polyamine synthesis [136], which could have a negative impact on cellular growth and differentiation during human development, since polyamine metabolism is critical to these processes [137].

Table 14.3 Metabolic pathways perturbed in hES cells following developmental toxicant exposure.

Alanine, aspartate, and glutamate metabolism	Members of this pathway are important for brain development and regulation of normal neurogenesis and apoptosis [124], as well as learning, memory, and cognition [125]
Arginine and proline metabolism	Includes observed elevations in asymmetric dimethylarginine (ADMA), which is an inhibitor of nitric oxide synthase (NOS). Nitric oxide (NO) formation is required for neural tube closure [126]
Citrate cycle (TCA cycle)	Metabolites in this pathway are key intermediates in energy production and impairments in this function result in neurological symptoms [127]
Cysteine and methionine metabolism	A broad pathway including cysteine, cystine, and cystathionine is involved in oxidative stress and other biochemical processes important for normal development (see Figure 14.2)
Glutathione metabolism	Decreased levels of the antioxidant glutathione cause limb malformations and embryopathy in animals [128,129] and are also related to neurodegenerative disease [130], pulmonary disease [131], and preeclampsia [132]
Nicotinate and nicotinamide metabolism	Nicotinate and nicotinamide are precursors of NAD^+ and $NADP^+$, which, when reduced, are important cofactors in many redox reactions. Mutations in nicotinamide *N*-methyl transferase (NNMT) have been associated with increased risk of spina bifida [133]
Pantothenate and CoA biosynthesis	Maternal pantothenate deficiency causes fetal resorptions, edema, cerebral, and eye defects [134,135]

In recent years, the role of cystine has been investigated with regard to its capacity to modulate cell growth, proliferation, apoptosis, and other cellular events. This has led to an understanding of cystine's role in numerous biological functions critical to human growth and development. Extracellularly, or within the secretome, cystine predominates over cysteine due to the oxidative state of the medium. Cystine is rapidly converted to cysteine once it is imported into the intracellular environment where it is a part of the cystine/cysteine thiol redox couple and used for many purposes, including the synthesis of glutathione, synthesis of proteins, and as a precursor for many other metabolites (Figure 14.2). The cystine/cysteine thiol redox couple is a critical component of a cell's regulatory capacity to handle reactive oxygen species (ROS). A broad spectrum of teratogens including pharmaceuticals, pesticides, and environmental contaminants are suspected of creating ROS or disrupting cellular mechanisms that maintain the appropriate balance of a cell's redox state, which can lead to adverse effects on developmental regulatory networks as a mechanism of action of developmental toxicity [138,139].

The predictivity and reproducibility of ornithine and cystine were initially evaluated using the untargeted metabolomic methods. Individually, ornithine and cystine were each able to classify the training set with 83% accuracy. When the metabolites were combined in a ratio of the ornithine fold change divided by the cystine fold change (i.e., the o/c ratio), the training set was predicted with

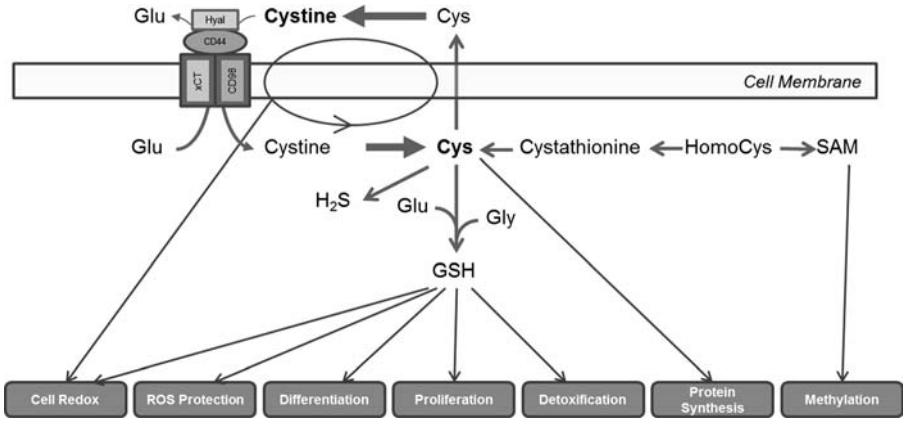


Figure 14.2 Cystine transporter mechanisms, cystine/cysteine redox, and the major cellular processes impacted by cystine transport and downstream metabolism. *Abbreviations*: Glu, glutamate; Cys, cysteine; HomoCys, homocysteine; Gly, glycine; SAM, *S*-adenosyl methionine; GSH, glutathione; ROS, reactive oxygen species; xCT, cystine/glutamate antiporter.

91% accuracy. A targeted metabolomic method was developed to measure specifically ornithine and cystine using a shorter data acquisition time in a single ionization mode, increasing instrument throughput eightfold. Since in toxicology it is a well-accepted premise that "the dose makes the poison," the targeted biomarker assay was developed using a nine-point dose–response curve to determine the exposure level at which a compound exhibits teratogenicity potential. The assay also includes a cell viability endpoint for determining whether changes in the o/c ratio are a direct result of cell death or due to metabolic changes unrelated to cytotoxicity. This novel assay was 77% accurate in identifying potential developmental toxicants in an independent test set. Two compounds that have been tested in the targeted biomarker assay are presented in Figure 14.3. Carbamazepine is a known human teratogen that is correctly predicted as a teratogen in the targeted biomarker assay (Figure 14.3a). The o/c ratio predicts carbamazepine to be teratogenic at 0.9 μM, a concentration well below the therapeutic C_{max}. In addition, carbamazepine elicits a response in the o/c ratio in the absence of cytotoxicity, demonstrating the sensitivity of a metabolic endpoint. Doxylamine, on the other hand, is nonteratogenic in humans. When hES cells are exposed to concentrations of doxylamine equivalent to the therapeutic C_{max}, there is no change in the o/c ratio (Figure 14.3b). However, at concentrations significantly higher than the C_{max}, doxylamine exposure causes a change in the o/c ratio indicative of teratogenicity, demonstrating the importance of incorporating the expected human exposure level in the prediction of toxicity. The targeted biomarker assay can help define exposure ranges where response may be expected as well as those where a response would not be expected to occur.

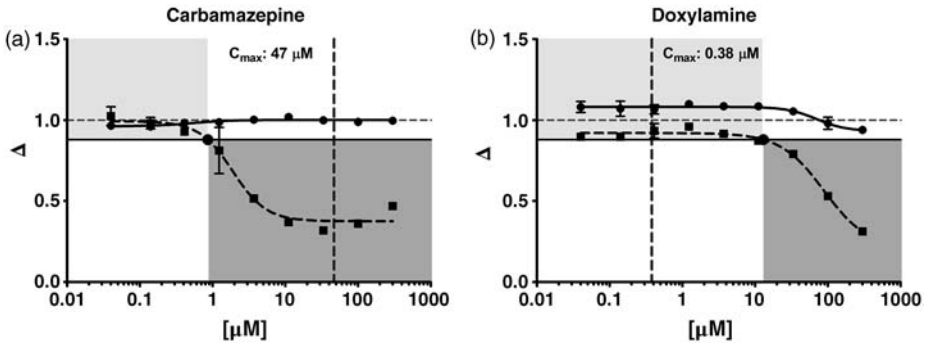


Figure 14.3 Illustration of results obtained using the targeted biomarker assay. The dose–response curves for the viability analysis (solid black curve, circle points) and o/c ratio (broken black curve, square points) are shown for the human teratogen carbamazepine (a) and human nonteratogen doxylamine (b). The *x*-axis is the concentration (μM) of the compound. Both the cell viability measurements and o/c ratio measurements exist on the same scale represented by Δ on the *y*-axis. The *y*-axis value of the o/c ratio is the ratio of the reference treatment normalized (fold change) values (ornithine/cystine). The *y*-axis value for the viability measurement is the treatment cell viability RFU normalized to the reference treatment cell viability RFU. The vertical broken dark gray line indicates the compound-specific C_{max} and the horizontal solid black line indicates the teratogenicity threshold (0.88). The filled black circle represents the concentration that a compound is predicted to have teratogenic potential. The light gray and dark gray shaded areas represent the concentrations where the compound is predicted to be nonteratogenic or teratogenic, respectively. The points are mean values and error bars are the standard error of the mean.

14.5.4
Drawbacks of *In Vitro* Models

There has been significant advancement in the development of *in vitro* biomarker-based assays for developmental toxicity, although as with other biological model systems, they cannot fully recapitulate the *in vivo* biology. The *in vitro* models do not include biology associated with the effects of absorption, distribution, metabolism, and excretion (ADME), which make it difficult to extrapolate doses, tissue/cellular compound delivery, and duration of exposure. Prior knowledge of compound metabolism may provide an opportunity to assess known metabolites in addition to the parent compound and provide information as to the relative toxicities of the metabolically related compounds, which is more difficult to ascertain using *in vivo* systems.

The "pathways of toxicity" concept suggests that validation of *in vitro* toxicity assays should include sufficient compounds to challenge all of the possible pathways. Without a map of all the relevant pathways, it is difficult to know whether models will be inclusive. Assays that include broad metabolic evaluation or rely on identified nodes of metabolism associated with toxicity such as the biomarker-based hES cell assay described here will be the most informative and predictive.

The uniqueness of using pluripotent stem cells addresses the biology of the early developing embryo; however, toxicities associated with differentiation and organogenesis may not be captured well. Additional aspects of maternal–fetal interactions, environment, genetics, and prenatal care (nutrition, drug use, etc.) also go beyond the scope of current *in vitro* assays. However, one of the advantages of using an *in vitro* assay is the ability to separate adverse outcomes due to compound from outcomes due to maternal toxicity from compound exposure.

Understanding the full potential and risks of a compound to induce developmental toxicity is important to efficiently moving compounds down the development path and the use of *in vitro* assays as screens with high throughput, reasonable cost, and reduced animal usage is an opportunity that should not be overlooked. It is also recognized that the current focus should be on replacing toxicity testing procedures with panels of predictive tests that can be done rapidly and efficiently [140,141]. Given the physiological relevance of hES cells to human development, developmental toxicity testing in cells derived from human embryos is likely to generate more reliable *in vitro* prediction endpoints than endpoints currently available through the use of animal models or other *in vitro* nonhuman assays.

14.6 Summary

A developmental toxicity assay that uses hPS cells and metabolomics, as described here, was highly predictive in identifying potential developmental toxicants with several advantages over other assays. The untargeted metabolomic approach allows the researcher to infer mechanistic information based on metabolite and pathway perturbations, while a targeted method focusing on specific biomarkers enables high-throughput capability. The use of a metabolomic approach gives the researcher two options for compound evaluation that can be used independently (or combined) for developmental toxicity prediction and aiding in compound prioritization: (1) using targeted biomarkers for a fast turnaround, simple prediction or (2) elucidating mechanisms or pathways of toxicity by surveying a wide breadth of metabolite changes with an untargeted approach. In addition, these cell-based assays offer a lower cost, more rapid turnaround, and reduced compound requirement compared with prior methods. This assay can help reduce or eliminate species-specific misinterpretations, since human cells are used, reduce the need for a second species in developmental toxicity testing, and could be included as part of a panel of *in vitro* assays aimed at defining where potential adverse responses in human populations may exist. Using an untargeted metabolomic endpoint, there is an opportunity to identify potential pathways of toxicology, which is a key concept in the Tox21c report published by the NRC [17]. Another key concept is to develop assays that can determine chemical concentration ranges where a toxic response may be observed. This question can begin to be addressed by designing assays that use a multiexposure

approach, as was done when the targeted biomarker assay was developed. The ongoing initiatives to move away from animal testing and develop *in vitro* models relevant to human biology are especially important in light of the Tox21 and REACH programs in the United States and Europe. The technologies and assays described in this chapter have the potential to address these needs.

References

1 Ujházy, E., Mach, M., Navarová, J., Brucknerová, I., and Dubovický, M. (2012) Teratology – past, present and future. *Interdiscip. Toxicol.*, **5**, 163–168.

2 Rogers, J.M. and Kavlock, R.J. (2007) Developmental toxicology, in *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 7th edn (ed. C.D. Klaassen), McGraw-Hill Professional.

3 Friedman, J.M. (2010) The principles of teratology: are they still true? *Birth Defects Res. A Clin. Mol. Teratol.*, **88**, 766–768.

4 Schwetz, B.A. and Harris, M.W. (1993) Developmental toxicology: status of the field and contribution of the National Toxicology Program. *Environ. Health Perspect.*, **100**, 269–282.

5 National Research Council (2000) *Scientific Frontiers in Developmental Toxicology and Risk Assessment*, National Academies Press, Washington, DC.

6 Centers for Disease Control and Prevention (2008) Update on overall prevalence of major birth defects – Atlanta, Georgia, 1978–2005. *MMWR Surveill. Summ.*, **57**, 1–5.

7 Christianson, A., Howson, C.P., and Modell, B. (2006) *The March of Dimes Global Report on Birth Defects: The Hidden Toll of Dying and Disabled Children*, March of Dimes Birth Defects Foundation, White Plains, NY.

8 Hamilton, B.E., Hoyert, D.L., Martin, J.A., Strobino, D.M., and Guyer, B. (2013) Annual summary of vital statistics: 2010–2011. *Pediatrics*, **131**, 548–558.

9 Petrini, J., Karla, D., and Johnston, R.B., Jr. (1997) An overview of infant mortality and birth defects in the United States. *Teratology*, **56**, 8–10.

10 Brent, R.L. and Beckman, D.A. (1990) Environmental teratogens. *Bull. N. Y. Acad. Med.*, **66**, 123–163.

11 Schardein, J.L. and Macina, O.T. (eds) (2007) *Human Developmental Toxicants: Aspects of Toxicology and Chemistry*, CRC Press, Boca Raton, FL.

12 Schardein, J.L. (ed.) (2000) *Chemically Induced Birth Defects*, 3rd edn, Marcel Dekker, New York.

13 Shepard, T. and Lemire, R. (eds) (2007) *Catalog of Teratogenic Agents*, 12th edn, The Johns Hopkins University Press, Baltimore, MD.

14 Pellizzer, C., Bremer, S., and Hartung, T. (2005) Developmental toxicity testing from animal towards embryonic stem cells. *ALTEX*, **22**, 47–57.

15 European Commission (2010) Sixth Report on the Statistics on the Number of Animals Used for Experimental and Other Scientific Purposes in the Member States of the European Union. Report from the Commission to the Council and the European Parliament, COM (2010) 511, Brussels.

16 Daston, G.P. and Knudsen, T.B. (2010) Fundamental concepts, current regulatory design and interpretation, in *Comprehensive Toxicology*, 2nd edn, vol. **12** (eds G. Daston and T. Knudsen), Elsevier, New York.

17 National Research Council (2007) *Toxicity Testing in the 21st Century: A Vision and a Strategy*, National Academies Press, Washington, DC.

18 Andrews, P.W. (2002) From teratocarcinomas to embryonic stem cells. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **357**, 405–417.

19 Wobus, A.M. and Boheler, K.R. (2005) Embryonic stem cells: prospects for

developmental biology and cell therapy. *Physiol. Rev.*, **85**, 635–678.
20 Evans, M.J. and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature*, **292**, 154–156.
21 Martin, G.R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA*, **78**, 7634–7636.
22 Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998) Embryonic stem cell lines derived from human blastocysts. *Science*, **282**, 1145–1147.
23 He, J., Ma, Y., Lee, Y., Thomson, J.A., and Kamp, T.J. (2003) Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization. *Circ. Res.*, **93**, 32–39.
24 Reubinoff, B.E., Pera, M.F., and Fong, C.Y. (2000) Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nat. Biotechnol.*, **18**, 399–404.
25 Zhang, S.C., Wering, M., Duncan, I.D., Brustle, O., and Thomson, J.A. (2001) *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. *Nat. Biotechnol.*, **19**, 1129–1133.
26 Keirstead, H.S., Nistor, G., Bernal, G., Totoiu, M., Cloutier, F., Sharp, K., and Steward, O. (2005) Human embryonic stem cell-derived oligodendrocyte progenitor cell transplants remyelinate and restore locomotion after spinal cord injury. *J. Neurosci.*, **25**, 4694–4705.
27 Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
28 Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861–872.
29 Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.
30 Puri, M.C. and Nagy, A. (2012) Concise review: Embryonic stem cells versus induced pluripotent stem cells: the game is on. *Stem Cells*, **30**, 10–14.
31 Jia, F., Wilson, K.D., Sun, N., Gupta, D.M., Huang, M., Li, Z., Panetta, N.F., Chen, Z.Y., Robbins, R.C., Kay, M.A., Longaker, M.T., and Wu, J.C. (2010) A nonviral minicircle vector for deriving human iPS cells. *Nat. Methods*, **7**, 197–199.
32 Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008) Generation of mouse induced pluripotent stem cells without viral vectors. *Science*, **322**, 949–953.
33 Si-Tayeb, K., Noto, F.K., Sepac, A., Sedlic, F., Bosnjak, Z.J., Lough, J.W., and Duncan, S.A. (2010) Generation of human induced pluripotent stem cells by simple transient transfection of plasmid DNA encoding reprogramming factors. *BMC Dev. Biol.*, **10**, 81.
34 Yu, J., Hu, K., Smuga-otto, K., Tian, S., Stewart, R., Igor, I., and Thomson, J.A. (2009) Human induced pluripotent stem cells free of vector and transgene sequences. *Science*, **324**, 797–801.
35 Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., and Mohseni, P. (2009) Virus free induction of pluripotency and subsequent excision of reprogramming factors. *Nature*, **458**, 771–775.
36 Somers, A., Jean, J.C., Sommer, C.A., Omari, A., Ford, C.C., Mills, J.A., Ying, L., Sommer, A.G., Jean, J.M., Smith, B.W., Lafyatis, R., Demierre, M.F., Weiss, D.J., French, D.L., Gadue, P., Murphy, G.J., Mostoslavsky, G., and Kotton, D.N. (2010) Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells*, **28**, 1728–1740.
37 Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H.-K.,

and Nagy, A. (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*, **458**, 766–770.

38 Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., Siuzdak, G., Schöler, H.R., Duan, L., and Ding, S. (2009) Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell*, **4**, 381–384.

39 Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., and Kim, K.-S. (2009) Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell*, **4**, 472–476.

40 Thier, M., Münst, B., Mielke, S., and Edenhofer, F. (2012) Cellular reprogramming employing recombinant sox2 protein. *Stem Cells Int.*, **2012**, 549846.

41 Thier, M., Münst, B., and Edenhofer, F. (2010) Exploring refined conditions for reprogramming cells by recombinant Oct4 protein. *Int. J. Dev. Biol.*, **54**, 1713–1721.

42 Tang, Y., Lin, C.J., and Tian, X.C. (2011) Functionality and transduction condition evaluation of recombinant Klf4 for improved reprogramming of iPS cells. *Cell. Reprogram.*, **13**, 99–112.

43 Zhang, H., Ma, Y., Gu, J., Liao, B., Li, J., Wong, J., and Jin, Y. (2012) Reprogramming of somatic cells via TAT-mediated protein transduction of recombinant factors. *Biomaterials*, **33**, 5047–5055.

44 Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.-H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., Daley, G.Q., Brack, A.S., Collins, J.J., Cowan, C., Schlaeger, T.M., and Rossi, D.J. (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*, **7**, 618–630.

45 Angel, M. and Yanik, M.F. (2010) Innate immune suppression enables frequent transfection with RNA encoding reprogramming proteins. *PLoS One*, **5**, e11756.

46 Drews, K., Tavernier, G., Demeester, J., Lehrach, H., De Smedt, S.C., Rejman, J., and Adjaye, J. (2012) The cytotoxic and immunogenic hurdles associated with non-viral mRNA-mediated reprogramming of human fibroblasts. *Biomaterials*, **33**, 4059–4068.

47 Tavernier, G., Wolfrum, K., Demeester, J., De Smedt, S.C., Adjaye, J., and Rejman, J. (2012) Activation of pluripotency-associated genes in mouse embryonic fibroblasts by non-viral transfection with *in vitro*-derived mRNAs encoding Oct4, Sox2, Klf4 and cMyc. *Biomaterials*, **33**, 412–417.

48 Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., and Morrisey, E.E. (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*, **8**, 376–388.

49 Miyoshi, N., Ishii, H., Nagano, H., Haraguchi, N., Dewi, D.L., Kano, Y., Nishikawa, S., Tanemura, M., Mimori, K., Tanaka, F., Saito, T., Nishimura, J., Takemasa, I., Mizushima, T., Ikeda, M., Yamamoto, H., Sekimoto, M., Doki, Y., and Mori, M. (2011) Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell*, **8**, 633–638.

50 Desponts, C. and Ding, S. (2010) Using small molecules to improve generation of induced pluripotent stem cells from somatic cells. *Methods Mol. Biol.*, **636**, 207–218.

51 Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008) Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nat. Biotechnol.*, **26**, 795–797.

52 Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Schöler, H.R., and Ding, S. (2008) A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell*, **2**, 525–528.

53 Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., Huangfu, D., Akutsu, H., Liu, D.R., Rubin,

L.L., and Eggan, K. (2009) A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell*, **5**, 491–503.

54 Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., and Ding, S. (2009) A chemical platform for improved induction of human iPSCs. *Nat. Methods*, **6**, 805–808.

55 Lyssiotis, C.A., Foreman, R.K., Staerk, J., Garcia, M., Mathur, D., Markoulaki, S., Hanna, J., Lairson, L.L., Charette, B.D., Bouchez, L.C., Bollong, M., Kunick, C., Brinker, A., Cho, C.Y., Schultz, P.G., and Jaenisch, R. (2009) Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. *Proc. Natl. Acad. Sci. USA*, **106**, 8912–8917.

56 Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., and Ding, S. (2010) Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell*, **7**, 651–655.

57 Lee, C.H., Kim, J.-H., Lee, H.J., Jeon, K., Lim, H., Choi, H.Y., Lee, E.-R., Park, S.H., Park, J.-Y., Hong, S., Kim, S., and Cho, S.-G. (2011) The generation of iPS cells using non-viral magnetic nanoparticle based transfection. *Biomaterials*, **32**, 6683–6691.

58 Pasha, Z., Haider, H.K., and Ashraf, M. (2011) Efficient non-viral reprogramming of myoblasts to stemness with a single small molecule to generate cardiac progenitor cells. *PLoS One*, **6**, e23667.

59 Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., Campos, L.S., Rad, R., Guo, G., Zhang, S., Bradley, A., and Liu, P. (2011) Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc. Natl. Acad. Sci. USA*, **108**, 18283–18288.

60 Robinton, D.A. and Daley, G.Q. (2012) The promise of induced pluripotent stem cells in research and therapy. *Nature*, **481**, 295–305.

61 Hug, K. (2006) Therapeutic perspectives of human embryonic stem cell research versus the moral status of a human embryo – does one have to be compromised for the other? *Medicina*, **42**, 107–114.

62 Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G.G., and Kell, D.B. (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.*, **22**, 245–252.

63 Khoo, S.H.G. and Al-Rubeai, M. (2007) Metabolomics as a complementary tool in cell culture. *Biotechnol. Appl. Biochem.*, **47**, 71–84.

64 Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.

65 Nielsen, J. and Oliver, S. (2005) The next wave in metabolome analysis. *Trends Biotechnol.*, **23**, 544–546.

66 Mapelli, V., Olsson, L., and Nielsen, J. (2008) Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. *Trends Biotechnol.*, **26**, 490–497.

67 Kell, D.B., Brown, M., Davey, H.M., Dunn, W.B., Spasic, I., and Oliver, S.G. (2005) Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol.*, **3**, 557–565.

68 Schmidt, C.W. (2004) Metabolomics: what's happening downstream of DNA. *Environ. Health Perspect.*, **112**, A410–A415.

69 Clarke, C.J. and Haselden, J.N. (2008) Metabolic profiling as a tool for understanding mechanisms of toxicity. *Toxicol. Pathol.*, **36**, 140–147.

70 Rochfort, S. (2005) Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J. Nat. Prod.*, **68**, 1813–1820.

71 Kostiainen, R., Kotiaho, T., Kuuranne, T., and Auriola, S. (2003) Liquid chromatography/atmospheric pressure ionization-mass spectrometry in drug metabolism studies. *J. Mass Spectrom.*, **38**, 357–372.

72 Ma, S., Chowdhury, S.K., and Alton, K.B. (2006) Application of mass spectrometry for metabolite identification. *Curr. Drug Metab.*, **7**, 503–523.

73 Schlotterbeck, G., Ross, A., Dieterle, F., and Senn, H. (2006) Metabolic profiling technologies for biomarker discovery in biomedicine and drug development. *Pharmacogenomics*, **7**, 1055–1075.

74 Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K., Fraser, P.D., and Bramley, P.M. (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.*, **56**, 219–243.

75 Shulaev, V. (2006) Metabolomics technology and bioinformatics. *Brief. Bioinform.*, **7**, 128–139.

76 Fernie, A.R., Trethewey, R.N., Krotzky, A.J., and Willmitzer, L. (2004) Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.*, **5**, 763–769.

77 Chen, Y., Monshouwer, M., and Fitch, W. (2007) Analytical tools and approaches for metabolite identification in early drug discovery. *Pharm. Res.*, **24**, 248–257.

78 Prakash, C., Shaffer, C.L., and Nedderman, A. (2007) Analytical strategies for identifying drug metabolites. *Mass Spectrom. Rev.*, **26**, 340–369.

79 Roberts, L.D., Souza, A.L., Gersten, R.E., and Clish, C.B. (2012) Targeted metabolomics. *Curr. Protoc. Mol. Biol.*, **30**, 30.2.1–30.2.24.

80 Cech, N.B. and Enke, C.G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.*, **20**, 362–387.

81 Dudley, E., Yousef, M., Wang, Y., and Griffiths, W.J. (2010) Targeted metabolomics and mass spectrometry. *Adv. Protein Chem. Struct. Biol.*, **80**, 45–83.

82 Patti, G.J., Yanes, O., and Siuzdak, G. (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.*, **13**, 263–269.

83 Vinayavekhin, N. and Saghatelian, A. (2010) Untargeted metabolomics. *Curr. Protoc. Mol. Biol.*, **30**, 30.1.1–30.1.24.

84 Baker, M. (2011) Metabolomics: from small molecules to big ideas. *Nat. Methods*, **8**, 117–121.

85 Ramirez, T., Daneshian, M., Kamp, H., Bois, F.Y., Clench, M.R., Coen, M., Donley, B., Fischer, S.M., Ekman, D.R., Fabian, E., Guillou, C., Heuer, J., Hogberg, H.T., Jungnickel, H., Keun, H.C., Krennrich, G., Krupp, E., Luch, A., Noor, F., Peter, E., Riefke, B., Seymour, M., Skinner, N., Smirnova, L., Verheij, E., Wagner, S., Hartung, T., van Ravenzwaay, B., and Leist, M. (2013) Metabolomics in toxicology and preclinical research. *ALTEX*, **20**, 209–225.

86 Robertson, D.G., Watkins, P.B., and Reily, M.D. (2011) Metabolomics in toxicology: preclinical and clinical applications. *Toxicol. Sci.*, **120** (Suppl.), S146–S170.

87 Bouhifd, M., Hartung, T., Hogberg, H.T., Kleensang, A., and Zhao, L. (2013) Review: Toxicometabolomics. *J. Appl. Toxicol.*, **33**, 1365–1383.

88 Sumner, S., Snyder, R., Burgess, J., Myers, C., Tyl, R., Sloan, C., and Fennell, T. (2009) Metabolomics in the assessment of chemical-induced reproductive and developmental outcomes using non-invasive biological fluids: application to the study of butylbenzyl phthalate. *J. Appl. Toxicol.*, **29**, 703–714.

89 Cuperlović-Culf, M., Barnett, D.A., Culf, A.S., and Chute, I. (2010) Cell culture metabolomics: applications and future directions. *Drug Discov. Today*, **15**, 610–621.

90 Van Vliet, E., Morath, S., Eskes, C., Linge, J., Rappsilber, J., Honegger, P., Hartung, T., and Coecke, S. (2008) A novel *in vitro* metabolomics approach for neurotoxicity testing, proof of principle for methyl mercury chloride and caffeine. *Neurotoxicology*, **29**, 1–12.

91 Ellis, J.K., Athersuch, T.J., Cavill, R., Radford, R., Slattery, C., Jennings, P., McMorrow, T., Ryan, M.P., Ebbels, T.M., and Keun, H.C. (2011) Metabolic response to low-level toxicant exposure in a novel renal tubule epithelial cell system. *Mol. Biosyst.*, **7**, 247–257.

92 Ruiz-Aracama, A., Peijnenburg, A., Kleinjans, J., Jennen, D., van Delft, J., Hellfrisch, C., and Lommen, A. (2011) An

untargeted multi-technique metabolomics approach to studying intracellular metabolites of HepG2 cells exposed to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. *BMC Genomics*, **12**, 251.

93 Balcke, G.U., Kolle, S.N., Kamp, H., Bethan, B., Looser, R., Wagner, S., Landsiedel, R., and van Ravenzwaay, B. (2011) Linking energy metabolism to dysfunctions in mitochondrial respiration – a metabolomics *in vitro* approach. *Toxicol. Lett.*, **203**, 200–209.

94 Vulimiri, S.V., Misra, M., Hamm, J.T., Mitchell, M., and Berger, A. (2009) Effects of mainstream cigarette smoke on the global metabolome of human lung epithelial cells. *Chem. Res. Toxicol.*, **22**, 492–503.

95 Kilk, K., Mahlapuu, R., Soomets, U., and Langel, U. (2009) Analysis of *in vitro* toxicity of five cell-penetrating peptides by metabolic profiling. *Toxicology*, **265**, 87–95.

96 Spielmann, H., Pohl, I., Döring, B., Liebsch, M., and Moldenhauer, F. (1997) The embryonic stem cell test (EST), an *in vitro* embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicol.*, **10**, 119–127.

97 Panzica-Kelly, J.M., Brannen, K.C., Ma, Y., Zhang, C.X., Flint, O.P., Lehman-McKeeman, L.D., and Augustine-Rauch, K.A. (2013) Establishment of a molecular embryonic stem cell developmental toxicity assay. *Toxicol. Sci.*, **131**, 447–457.

98 Theunissen, P.T. and Piersma, A.H. (2012) Innovative approaches in the embryonic stem cell test (EST). *Front. Biosci. (Landmark Ed.)*, **17**, 1965–1975.

99 Goetz, A.K. and Dix, D.J. (2009) Mode of action for reproductive and hepatic toxicity inferred from a genomic study of triazole antifungals. *Toxicol. Sci.*, **110**, 449–462.

100 Jergil, M., Kultima, K., Gustafson, A.L., Dencker, L., and Stigson, M. (2009) Valproic acid-induced deregulation *in vitro* of genes associated *in vivo* with neural tube defects. *Toxicol. Sci.*, **108**, 132–148.

101 Bigot, K., de Lange, J., Archer, G., Clothier, R., and Bremer, S. (1999) The relative semi-quantification of mRNA expression as a useful toxicological endpoint for the identification of embryotoxic/teratogenic substances. *Toxicol. In Vitro*, **13**, 619–623.

102 Pellizzer, C., Adler, S., Corvi, R., Hartung, T., and Bremer, S. (2004) Monitoring of teratogenic effects *in vitro* by analysing a selected gene expression pattern. *Toxicol. In Vitro*, **18**, 325–335.

103 Van Dartel, D.A.M. and Piersma, A.H. (2011) The embryonic stem cell test combined with toxicogenomics as an alternative testing model for the assessment of developmental toxicity. *Reprod. Toxicol.*, **32**, 235–244.

104 Van Dartel, D.A.M., Pennings, J.L.A., Robinson, J.F., Kleinjans, J.C.S., and Piersma, A.H. (2011) Discriminating classes of developmental toxicants using gene expression profiling in the embryonic stem cell test. *Toxicol. Lett.*, **201**, 143–151.

105 Amit, M., Carpenter, M.K., Inokuma, M.S., Chiu, C.P., Waknitz, M.A., Itskovitz-Eldor, J., and Thomson, J.A. (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.*, **227**, 271–278.

106 Wobus, A.M. and Löser, P. (2011) Present state and future perspectives of using pluripotent stem cells in toxicology research. *Arch. Toxicol.*, **85**, 79–117.

107 Scott, C.W., Peters, M.F., and Dragan, Y.P. (2013) Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. *Toxicol. Lett.*, **219**, 49–58.

108 Adler, S., Pellizzer, C., Hareng, L., Hartung, T., and Bremer, S. (2008) First steps in establishing a developmental toxicity test method based on human embryonic stem cells. *Toxicol. In Vitro*, **22**, 200–211.

109 Adler, S., Lindqvist, J., Uddenberg, K., Hyllner, J., and Strehl, R. (2008) Testing potential developmental toxicants with a cytotoxicity assay based on human embryonic stem cells. *Altern. Lab. Anim.*, **36**, 129–140.

110 Behar, R.Z., Bahl, V., Wang, Y., Lin, S., Xu, N., Davis, B., and Talbot, P. (2012) A method for rapid dose-response screening of environmental chemicals using human embryonic stem cells. *J. Pharmacol. Toxicol. Methods*, **66**, 238–245.

111 Mehta, A., Konala, V.B.R., Khanna, A., and Majumdar, A.S. (2008) Assessment of drug induced developmental toxicity using human embryonic stem cells. *Cell Biol. Int.*, **32**, 1412–1424.

112 Flora, S.J. and Mehta, A. (2009) Monoisoamyl dimercaptosuccinic acid abrogates arsenic-induced developmental toxicity in human embryonic stem cell-derived embryoid bodies: comparison with *in vivo* studies. *Biochem. Pharmacol.*, **78**, 1340–1349.

113 Zdravkovic, T., Genbacev, O., LaRocque, N., McMaster, M., and Fisher, S. (2008) Human embryonic stem cells as a model system for studying the effects of smoke exposure on the embryo. *Reprod. Toxicol.*, **26**, 86–96.

114 Krishnamoorthy, M., Gerwe, B.A., Scharer, C.D., Heimburg-Molinaro, J., Gregory, F., Nash, R.J., Arumugham, J., Stewart, B., and Stice, S.L. (2010) Low ethanol concentration alters CHRNA5 RNA levels during early human development. *Reprod. Toxicol.*, **30**, 489–492.

115 Lin, S., Fonteno, S., Weng, J.H., and Talbot, P. (2010) Comparison of the toxicity of smoke from conventional and harm reduction cigarettes using human embryonic stem cells. *Toxicol. Sci.*, **118**, 202–212.

116 Wilson, K.D., Sun, N., Huang, M., Zhang, W.Y., Lee, A.S., Li, Z., Wang, S.X., and Wu, J.C. (2010) Effects of ionizing radiation on self-renewal and pluripotency of human embryonic stem cells. *Cancer Res.*, **70**, 5539–5548.

117 Pal, R., Mamidi, M.K., Das, A.K., and Bhonde, R. (2011) Human embryonic stem cell proliferation and differentiation as parameters to evaluate developmental toxicity. *J. Cell. Physiol.*, **226**, 1583–1595.

118 Zhu, M.-X., Zhao, J.-Y., Chen, G.-A., and Guan, L. (2011) Early embryonic sensitivity to cyclophosphamide in cardiac differentiation from human embryonic stem cells. *Cell Biol. Int.*, **35**, 927–938.

119 Cezar, G.G., Quam, J.A., Smith, A.M., Rosa, G.J.M., Piekarczyk, M.S., Brown, J.F., Gage, F.H., and Muotri, A.R. (2007) Identification of small molecules from human embryonic stem cells using metabolomics. *Stem Cells Dev.*, **16**, 869–882.

120 Palmer, J.A., Poenitzsch, A.M., Smith, S.M., Conard, K.R., West, P.R., and Cezar, G.G. (2012) Metabolic biomarkers of prenatal alcohol exposure in human embryonic stem cell-derived neural lineages. *Alcohol. Clin. Exp. Res.*, **36**, 1314–1324.

121 Kleinstreuer, N.C., Smith, A.M., West, P.R., Conard, K.R., Fontaine, B.R., Weir-Hauptman, A.M., Palmer, J.A., Knudsen, T.B., Dix, D.J., Donley, E.L.R., and Cezar, G.G. (2011) Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.*, **257**, 111–121.

122 West, P.R., Weir, A.M., Smith, A.M., Donley, E.L.R., and Cezar, G.G. (2010) Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.*, **247**, 18–27.

123 Palmer, J.A., Smith, A.M., Egnash, L.a., Conard, K.R., West, P.R., Burrier, R.E., Donley, E.L.R., and Kirchner, F.R. (2013) Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. *Birth Defects Res. B Dev. Reprod. Toxicol.*, **98**, 343–363.

124 Kemper, T.L. and Bauman, K.A. (1998) Neuropathology of infantile autism. *J. Neuropathol. Exp. Neurol.*, **57**, 645–652.

125 Tashiro, A., Sandler, V.M., Toni, N., Zhao, C., and Gage, F.H. (2006) NMDA-receptor-mediated, cell-specific integration of new neurons in adult dentate gyrus. *Nature*, **442**, 929–933.

126 Nachmany, A., Gold, V., Tsur, A., Arad, D., and Weil, M. (2006) Neural tube closure depends on nitric oxide synthase activity. *J. Neurochem.*, **96**, 247–253.

127 Rock, C.O., Calder, R.B., Karim, M.A., and Jackowski, S. (2000) Pantothenate kinase regulation of the intracellular concentration of coenzyme A. *J. Biol. Chem.*, **275**, 1377–1383.

128 Hansen, J.M., Carney, E.W., and Harris, C. (2001) Altered differentiation in rat and rabbit limb bud micromass cultures by glutathione modulating agents. *Free Radic. Biol. Med.*, **31**, 1582–1592.

129 Isibashi, M., Akazawa, S., Sakamake, H., Matsumoto, K., Yamasaki, H., Yamaguch, Y., Goto, S., Urata, Y., Kondo, T., and Nagataki, S. (1997) Oxygen-induced embryopathy and the significance of glutathione-dependent antioxidant system in the rat embryo during early organogenesis. *Free Radic. Biol. Med.*, **22**, 447–454.

130 Simonian, N.A. and Coyle, J.T. (1996) Oxidative stress in neurodegenerative diseases. *Annu. Rev. Pharmacol. Toxicol.*, **36**, 83–106.

131 Repine, J.E., Bast, A., and Lankhorst, I. (1997) Oxidative stress in chronic obstructive pulmonary disease. Oxidative Stress Study Group. *Am. J. Respir. Crit. Care Med.*, **156**, 341–357.

132 Walsh, S.W. (1998) Maternal–placental interactions of oxidative stress and antioxidants in preeclampsia. *Semin. Reprod. Endocrinol.*, **16**, 93–104.

133 Lu, W., Zhu, H., Wen, S., Yang, W., SHaw, G., Lammer, E., and Finnell, R. (2008) Nicotinamide *N*-methyl transferase (NNMT) gene polymorphisms and risk for spina bifida. *Birth Defects Res. C Embryo Today Rev.*, **82**, 670–675.

134 Baker, H., Frank, O., Thomson, A.D., Langer, A., Munves, E.D., De Angelis, B., and Kaminetzky, H.A. (1975) Vitamin profile of 174 mothers and newborns at parturition. *Am. J. Clin. Nutr.*, **28**, 59–65.

135 Nelson, M.M., Wright, H.V., Baird, C.D., and Evans, H.M. (1957) Teratogenic effects of pantothenic acid deficiency in the rat. *J. Nutr.*, **62**, 395–405.

136 Mao, Y., Gurr, J.A., and Hickok, N.J. (1993) Retinoic acid regulates ornithine decarboxylase gene expression at the transcriptional level. *Biochem. J.*, **295**, 641–644.

137 Kalhan, S.C. and Bier, D.M. (2008) Protein and amino acid metabolism in the human newborn. *Annu. Rev. Nutr.*, **28**, 389–410.

138 Hansen, J.M. (2006) Oxidative stress as a mechanism of teratogenesis. *Birth Defects Res. C Embryo Today Rev.*, **78**, 293–307.

139 Kovacic, P. and Somanathan, R. (2006) Mechanism of teratogenesis: electron transfer, reactive oxygen species, and antioxidants. *Birth Defects Res. C Embryo Today Rev.*, **78**, 308–325.

140 Andersen, M.E. and Krewski, D. (2009) Toxicity testing in the 21st century: bringing the vision to life. *Toxicol. Sci.*, **107**, 324–330.

141 Bhattacharya, S., Zhang, Q., Carmichael, P.L., Boekelheide, K., and Andersen, M.E. (2011) Toxicity testing in the 21st century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *PLoS One*, **6**, e20887.

15
Immunogenicity of Protein Therapeutics: Risk Assessment and Risk Mitigation

Harald Kropshofer

15.1
Introduction

In 1982, insulin was the first therapeutic protein that was approved for the treatment of humans. Since then, more than 150 therapeutic proteins have entered the marketplace worldwide [1,2]. Therapeutic biologics offer the advantages of increased specificity and reduced toxicity compared with small molecules. However, while small molecules easily escape recognition by the human immune system, protein therapeutics often surpass the thresholds for activating T and B lymphocytes, thereby becoming immunogenic.

Currently, protein therapeutics hold an expanding place among medicinal products: recombinant growth factors, cytokines, enzyme replacement proteins, monoclonal antibodies, and fusion proteins have been approved for use in the treatment of human diseases. As a consequence of the increasing number of protein therapeutics under development, adverse effects associated with immune responses to protein therapies became a subject of concern. For example, anti-therapeutic antibodies (ATAs) that develop in response to a therapeutic protein may alter the drug's pharmacokinetic profile and abrogate its pharmacodynamic effect (neutralizing activity) [3,4] or may cause safety concerns, such as hypersensitivity reactions [5]. ATA formation may rely on activated $CD4^+$ T cells or may be independent of them. T-cell-independent ATA responses may be generated when B cells recognize a repeated pattern (motif) of the therapeutic protein. Those B cells may respond by transiently producing low-affinity, predominantly IgM antibodies [6]. Antibodies that are generated in conjunction with T-cell help are referred to as T-cell-dependent or thymus-dependent ATAs. The latter process involves a complex interplay among antigen-presenting cells (APCs), T cells, secreted cytokines, and B cells. Hence, the presence of the IgG type of ATAs indicates that T cells are involved in the immune response to a therapeutic protein. Moreover, a number of clinical studies now suggest that high levels of T-cell-driven IgG ATAs have the potential to cross-react with the endogenous counterpart, an adverse effect that can have serious consequences [3–6]. These types of serious

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

outcomes resulting from cross-reactive ATAs have inspired the development of a multitude of *in vitro* methods for measuring the presence of ATAs, which have been carefully reviewed in several white papers [7–10]. In addition, methods for identifying drivers of T-cell-dependent immune responses have also expanded and have been described in a number of publications in the past decade [11–14].

Most recently, the European Medicines Agency (EMA) has published a "Guideline on Immunogenicity Assessment of Biotechnology-Derived Therapeutic Proteins" [15], in which factors influencing the immunogenicity of therapeutic proteins were classified into patient-, disease-, or product-related categories. The patient- and disease-related categories describe factors that may predispose a particular individual to an undesired immune response. In contrast, product-related factors, that is, factors intrinsic to the final drug product itself that contribute to immunogenicity, may include modifications in the glycosylation profile, biophysical and biochemical attributes, or factors introduced during formulation [16–19]. Table 15.1 summarizes some of these patient-, disease-, and product-related factors that have the potential to influence immune responses to a biological therapeutic. Many of these factors

Table 15.1 Aspects increasing the immunogenic risk of a therapeutic protein.

Context	Aspect	Rationale
Patient	Genotype	• HLA haplotype: binder of T-cell epitopes • Predisposition to allergy
	Other medication	• Premedication giving rise to cross-reactive ATA
	Others	• Age and sex • Pre-existing ATA
Disease	Immune status	• TH2 status in allergic diseases • Proinflammation in autoimmunity
	Genotype	• Expression of susceptibility factors, for example, particular HLA haplotypes
Product	Drug substance	• Sequence: neo T-cell epitopes • Nonhuman glycosylation pattern • Formation of aggregates *in situ* • Mimic of nonredundant endogenous protein
	Delivery	• Intermittent dosing • s.c. injection: erroneous exposure to dermis
	Formulation	• Excipients with costimulatory potential • Extreme pH difference to neutrality • Chelating agents
	Manufacturing + storage	• Contamination with host cell proteins • Manufacturing side products • Impurities, leachates (e.g., silicone, etc.) • Oxidation, deamidation, isomerization, and so on
	Target	• Receptor on immune cells

have been identified as "critical quality attributes" in the FDA-sponsored Quality-by-Design initiative that focuses on manufacturing "process development" [20].

To better control the immunogenic potential of protein therapeutics, more attempts are required to predict, mitigate, and monitor unwanted immune responses during both the preclinical and the clinical development, in particular in those cases where immune responses are suspected to give rise to safety issues. This chapter will discuss current approaches and the strategies of how to mitigate the immunogenicity risk in the clinical development of next-generation protein therapeutics.

15.2 The Central Role of CD4+ T Cells

Typical immune responses to therapeutic proteins can be divided into two categories: (i) activation of the classical adaptive immune system by so-called foreign proteins, such as immune responses against pathogens or vaccines; therapeutic proteins that replace endogenously lacking proteins would be viewed as "foreign" to the immune system of an individual who is missing, in whole or in part, the endogenous counterpart; and (ii) breakage of B- and/or T-cell tolerance, for example, the response elicited to self-proteins in certain autoimmune diseases. The underlying mechanisms are less well defined, but may include epitope mimicry, cross-reactive T cells, and presence of activators of the innate immune system, such as toll-like receptor agonists and/or aggregated proteins.

With regard to the classical immune pathway, production of ATAs is the culmination of a series of events that leads to B-cell activation and subsequently to differentiation into plasma cells that secrete ATAs. ATAs may be generated under the control of $CD4^+$ T cells or independent of T cells [21]. B cells are activated in a T-cell-independent manner when particular structural patterns, such as repeats of structural elements or carbohydrate molecules, directly induce activation of B cells by cross-linking of the B-cell receptor. The resulting ATA response is limited in both isotype (mainly IgM) and affinity; memory B cells are either not generated or short-lived [21]. In contrast, $CD4^+$ T-cell-dependent activation of B cells is characterized by class switching (IgM to IgG) and development of memory B cells that produce high-affinity and long-lived ATA responses. The development of ATAs of the IgG type following the administration of a therapeutic protein generally indicates that the therapeutic is driving a T-cell-dependent immune response.

Likewise, $CD4^+$ T-cell help is required for the class switch from IgG ATA to IgE ATA responses, which result in hypersensitivity reactions, such as type I allergy, mediated by cross-linking of drug-specific IgE on the surface of mast cells and/or basophils (Figure 15.1).

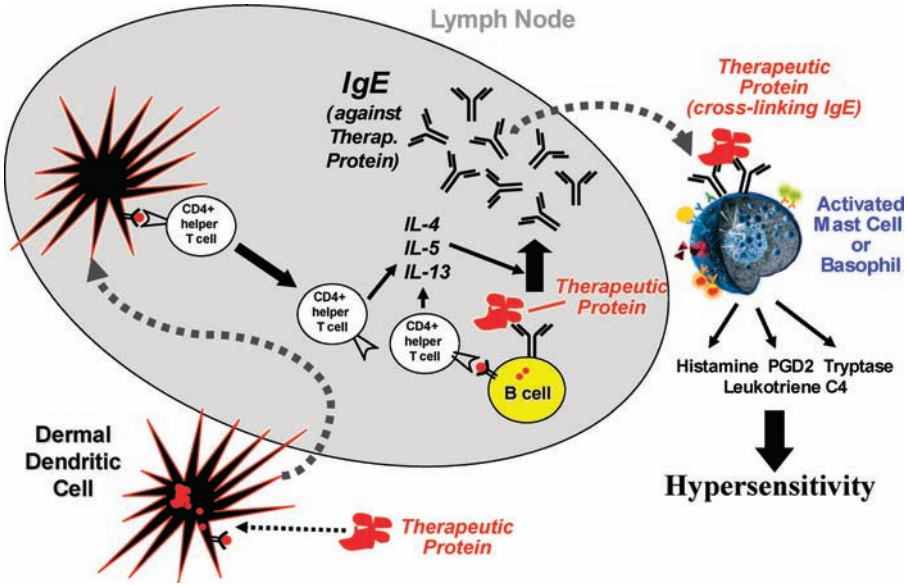


Figure 15.1 The cellular orchestra giving rise to hypersensitivity reactions triggered by therapeutic proteins. A therapeutic protein that is injected subcutaneously may have access to dermal dendritic cells that are able to endocytose and cleave the protein into peptides. Peptides may form HLA-DR–peptide complexes that – upon movement of the respective dendritic cells into dermal lymph nodes – may activate $CD4^+$ T cells. Interaction of these activated T cells with the same HLA-DR–peptide complexes on B cells may trigger B-cell differentiation and production of anti-drug antibodies (ATA). ATA generation may, eventually, switch from the initial IgG to IgE production. IgEs bound by Fcε receptors on mast cells or basophils may be cross-linked by the therapeutic protein they are specific for, leading to secretion of histamine, tryptase, prostaglandin D2 (PD2), and leukotriene C4, which give rise to a hypersensitivity reaction, for example, type I allergy or anaphylaxis.

15.3 Generation of T-Cell Epitopes

15.3.1 HLA Restriction

$CD4^+$ T-cell-dependent ATA responses, by definition, are restricted by the recognition of therapeutic protein-derived linear peptide fragments in the context of human leukocyte antigen (HLA) class II molecules. These peptide fragments are denoted as "T-cell epitopes" and are generated by APCs in the course of protein antigen processing [22]. To this end, the therapeutic protein is taken up by APCs and proteolytically processed into small peptides. Peptides that have a length of at least 13 amino acids and a motif of so-called anchor amino acid side chains, which fit into dedicated specificity pockets in the sole HLA peptide

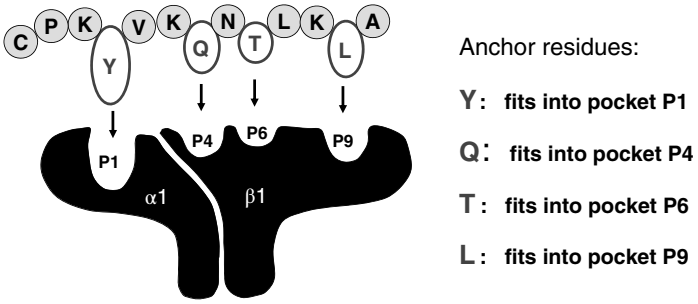


Figure 15.2 Binding of a T-cell epitope to a HLA-DR molecule. Sequence of the immunodominant influenza virus hemagglutinin T-cell epitope that is promiscuously binding to many allotypes of HLA-DR molecules (as shown by the two subunits α1 and β1). The HLA-DR molecule carries four specificity pockets, denoted as P1, P4, P6, and P9, which accommodate side chains of the peptide epitope, denoted as "anchor residues": Y, Q, T, and L fit into pockets P1, P4, P6, and P9, respectively. The residues up- and downstream of the anchor residues are potential contact sites for the interaction with the T-cell receptor.

binding cleft, may form HLA–peptide complexes (Figure 15.2). The appropriate set of anchor residues is a guarantee for the formation of high-stability HLA–peptide complexes, thereby facilitating sustained presentation of the respective HLA–peptide complexes on the APC cell surface in order to activate $CD4^+$ T cells [22]. As human populations express a huge variety of different allelic variants of HLA class II molecules, each peptide may bind to more than one HLA allelic variant and, likewise, a single HLA allotype may form complexes with more than one candidate T-cell epitope. The genetic polymorphism of the HLA class II loci and its impact on the binding of specific peptide ligands make the HLA genotype of patients a major determinant controlling immune responses to protein therapeutics.

$CD4^+$ T-cell recognition of a specific HLA–peptide complex, combined with costimulatory signals delivered by the APCs, culminates in robust T-cell activation that, in turn, facilitates a mature ATA response. In the absence of activated T helper cells, naïve B cells do not fully mature, and activated antigen-specific B cells are rendered anergic or undergo apoptosis. Therefore, T-cell recognition of peptide epitopes derived from protein therapeutics is a key determinant of $CD4^+$ T-cell-dependent ATA formation.

A typical T-cell-dependent ATA response to a therapeutic protein antigen can occur after one or two administrations of the respective protein therapeutic, often persists for prolonged periods of time, and ultimately may neutralize the therapeutic protein. Such an immune response is underlying the induction of ATAs directed against blood replacement factor VIII (FVIII) in hemophilia patients. Although the incidence and intensity of the immune response to FVIII can vary depending on the extent to which endogenous FVIII is expressed in the individual patient, ATA responses to FVIII are controlled by particular T-cell epitopes [23].

15.3.2
T-Cell Epitopes Controlling Immunogenicity

The presence versus abundance and the number of T-cell epitopes of a therapeutic protein can be exploited, in a first approximation, to assess the immunogenic risk. The underlying assumption is that the more stable the HLA–peptide complexes formed the higher the likelihood that naïve $CD4^+$ T cells will be activated and, subsequently, the more likely it is that B cells will generate ATAs [24]. This assumption in most instances leads to overprediction of immunogenicity, as several aspects that are not taken into account by the T-cell prediction approach may limit the formation of ATAs: (i) some T-cell epitopes may not give rise to HLA–peptide complexes as they are never formed during antigen processing (e.g., due to lack of protease cleavage sites), or they are formed but subsequently destroyed by proteolysis; (ii) some T-cell epitopes may be removed from the HLA binding cleft by the peptide editor HLA-DM as a consequence of kinetic proofreading [24]; (iii) some HLA–peptide complexes may not activate $CD4^+$ T cells due to lack of costimulatory signals, lack of appropriate T-cell receptors, or due to peripheral T-cell tolerance; and (iv) some HLA–peptide complexes may activate $CD4^+$ T cells that fail in activating B cells due to lack of costimulation, lack of appropriate B cells, or due to B-cell tolerance. Hence, the localization of T-cell epitopes within therapeutic protein sequences may be relevant as a first estimation of the immunogenic potential of a protein therapeutic [16,18,24]. Other considerations may have to follow in order to achieve a more meaningful assessment. Importantly, the contribution of regulatory T cells (Tregs), which may produce inhibitory signals in response to particular T-cell epitopes, has gained appreciation in the recent past. Methods for discriminating between regulatory and helper T-cell responses (Treg and T helper, respectively) are discussed below.

15.4
Tolerance to Therapeutic Drugs

The absence of an immune response to autologous proteins is attributed to central tolerance to proteins of an individual's proteome. This is due to the fact that during thymic development, T cells that respond to epitopes derived from autologous proteins expressed in the thymus undergo deletion or are rendered anergic [25]. However, tolerance to autologous proteins is incomplete: autoreactive T cells do exist in the periphery in the context of autoimmunity and are also present in the circulation of healthy individuals. Natural regulatory T cells (nTregs) are generated in the thymus and circulate in the periphery. Upon activation, $CD4^+CD25^+FoxP3^+$ nTregs are able to suppress bystander effector T cells directed against unrelated antigens [26,27]. It is tempting to assume that nTregs also control immune responses to therapeutic proteins that mimic or are identical to autologous proteins, such as erythropoietin, thrombopoietin, or glucagon-like peptide.

Adaptive Tregs, developed in reaction to foreign antigens to which central tolerance may not exist, are an additional source of control over immune responses [25]. Sustained tolerance (to exogenous proteins) may require the existence of these "adaptive" or "induced" Treg cells (aTreg, also known as iTreg) [26], with the same antigen specificity as the self-reactive effector T cells. Administration of antigen in the absence of an innate immune stimulator (danger signal) can lead to tolerance; this approach has been used for the induction of tolerance to allergens. A strong link connecting HLA, presentation of T-cell epitopes (both regulatory and effector) in the context of HLA, and the maintenance of peripheral tolerance has been described [28]. The implications of immune control by nTregs and the induction of tolerance via aTregs will be discussed in greater detail below.

15.5 Tool Set for Immunogenicity Risk Assessment

Given the complexity of aspects that apparently control T-cell-dependent ATA responses against protein therapeutics, immunogenicity assessment involves a multistep approach. The first step may be to analyze the amino acid sequence of the protein of interest for the presence of T-cell epitopes. Subsequent steps may exploit a variety of human cell-based methodologies, including HLA/peptide binding assays, assays based on whole blood-derived cells, or humanized mouse models.

15.5.1 Epitope Determination

15.5.1.1 *In Silico* Screening

While the medium length of HLA-associated T-cell epitopes is 15–16 amino acids, the core that mainly defines the stability of HLA–peptide complexes has a length of only 9–10 amino acids [29]. As a consequence, algorithms localizing potential T-cell epitopes based on the amino acid sequence of a protein have been developed. Databases such as the Immune Epitope Database Analysis Resource (IEDB; www.tools.immunoepitope.org) provide the raw material for developing T-cell epitope prediction tools [30]. A common denominator among these tools is the ability to quickly screen large data sets, including whole genomes or proteomes, for putative T-cell epitopes. The vast majority of *in silico* tools used in pharmaceutical development focus on T-cell epitopes restricted by HLA-DR molecules, mainly because there is no evidence that the other two types of HLA class II molecules, HLA-DQ or HLA-DP, contribute substantially to the immunogenicity of therapeutic proteins.

The computational approach to localize helper T-cell epitopes is well accepted in vaccine discovery efforts [31] and in approaches to identify epitopes that contribute to autoimmunity [32]. In the context of therapeutic proteins, several

studies have demonstrated the predictive potential of these *in silico* algorithms [33,34]. In parallel, tools have been developed to rank therapeutic proteins according to their T-cell epitope content [35] and to define means of modifying the epitope content of therapeutic proteins [36]. Application of *in silico* tools reduces downstream *in vitro* testing and provides an opportunity in early development to decrease the immunogenic potential of protein therapeutics. As the currently available *in silico* algorithms do not take into account proteolytic destruction of potential T-cell epitopes, competition at the HLA level by endogenous peptides, or failures in the cellular generation of T-cell epitopes, *in silico* screening performed in isolation is known to be overpredictive, leading to false positives. However, *in silico* tools provide the advantages of high throughput, low cost, and the ability to reduce the search space for downstream testing, such as HLA binding, antigen processing, and T-cell assays that complement *in silico* testing.

15.5.1.2 Peptide Elution

An assay that overcomes the issue of overprediction of *in silico* tools is the MAPPs (MHC-associated peptide proteomics) technology [29]. It is based on human dendritic cells that are coincubated with the therapeutic protein of choice *in vitro* and subsequent mass spectrometry-based sequencing of peptides after elution from HLA-DR molecules [37]. This technology allows the identification of truly processed and presented peptide epitopes derived from the cellular source, human dendritic cells, which are responsible for priming of $CD4^+$ T cells *in vivo*. The MAPPs approach led to the identification of novel helper T-cell epitopes in the context of tumor antigens [38] and T-cell epitopes of the therapeutic protein Betaseron, which are supposed to give rise to neutralizing ATA responses (unpublished results). While the MAPPs technology allows derisking those T-cell epitopes that are predicted by *in silico* tools rather than being presented to T cells, there is still room for improvement with regard to sample throughput and cost, thereby making this approach less attractive for screening purposes.

15.5.2 HLA Binding Assays

In vitro HLA class II binding assays can be used to confirm the T-cell epitopes located by *in silico* screening and peptide elution. Different test principles are available, as outlined in the following.

15.5.2.1 Competition Binding Assay

Here, peptides supposed to be T-cell epitopes are tested for their ability to compete against a labeled peptide that is known to be a cognate T-cell epitope in association with the respective HLA class II molecule [39]. A nonlinear regression analysis is used to calculate the IC_{50} (concentration of test peptide required to compete off 50% of the target peptide). Binding assays can be

performed for a broad representation of HLA alleles to cover the ethnic groups of interest [39]. One drawback of this method is that the resulting affinity values are relative to a standard reference peptide. As the reference peptides differ from HLA allele to allele, the outcome of this analysis is relative affinities and a true comparison is valid only for the same HLA allele but not across HLA alleles.

15.5.2.2 Real-Time Kinetic Measurements

This technology determines the rate at which peptides interact with HLA molecules *in vitro*. Protocols based on fluorescence polarization are suitable for kinetic studies, allowing multiple readings to be taken throughout the binding reaction [40]. In addition, surface plasmon resonance methods may be employed to measure peptide binding to HLA class II molecules in real time [41].

In summary, HLA binding assays may improve the accuracy of immunogenicity predictions when applied in a stepwise process after *in silico* screening of whole protein sequences and before the conduct of a biological assay, such as enzyme-linked immunosorbent spot-forming (ELISpot) assays or a T-cell activation assay.

15.5.3 T-Cell Activation Assays

In vitro assays involving peripheral blood-derived T cells have been used in transplantation research to assess the risk of engrafted T cells reacting against host tissue (graft-versus-host disease (GvHD)). Adaptation of these assays to the development of protein therapeutics adds another level of scrutiny to preclinical assessment of the risk of immunogenicity. Several surrogate markers for T-cell activation have been established for these *in vitro* assays, such as cytokine secretion, surface markers of activation, signal transduction events, and proliferation.

15.5.3.1 Cytokine Release

Enzyme-linked immunosorbent assay (ELISA) and ELISpot are two related methods for measuring cytokines secreted by activated T cells (i.e., IFN-γ, IL-2, IL-4, and IL-10) [42]. The ELISA is used to qualitatively and quantitatively analyze the cytokines in culture supernatants generated under conditions of T-cell stimulation. The ELISpot assay provides information regarding the number of cytokine-producing cells (down to one cell per million) within a cell population stimulated *ex vivo*. The ELISpot assay is considered to be more sensitive and quantitative than the ELISA.

Intracellular cytokine staining measured by flow cytometry is another method for detecting cytokines and linking their expression to the phenotype of individual cells [43]. The flow cytometry test can be used to accurately measure T-cell functionality relative to the phenotypic classification of $CD4^+$ T cells based on cell surface markers.

15.5.3.2 T-Cell Proliferation

T-cell proliferation in response to stimulation by a HLA–peptide complex can be measured either by the incorporation of tritiated thymidine into the DNA of dividing cells or by the dilution of a fluorescent dye, carboxyfluorescein succinimidyl ester (CFSE), that decreases in fluorescence intensity by half with each round of cell division, as determined by flow cytometry. In addition to CFSE labeling, cells can be costained for expression of markers that allow differentiation of regulatory versus (Th1/Th2) effector T-cell phenotypes.

To address the HLA diversity of a given patient population, the number of individual blood samples that need to be tested is quite large ($n > 40$), and the blood volume per sample ranges from 15 ml to more than 50 ml. Hence, maintenance of a large supply of blood samples from prequalified donors is mandatory.

15.5.3.3 Tetramers

Fluorescently labeled tetrameric complexes of HLA class II molecules loaded with the peptide of interest (i.e., "tetramers") can be used to quantitatively analyze T cells recognizing a particular T-cell epitope.

15.5.3.4 Naïve T-Cell Assay

Naïve peripheral blood mononuclear cells (PBMCs) are being used to assess the immunogenicity potential of therapeutic proteins [11,14,44]. It was postulated that the higher the frequency of precursor T cells that recognize a certain T-cell epitope, the higher the immunogenic potential of the respective T-cell epitope [45]. Antigen stimulation, sometimes over several weeks, is often required to expand sufficient numbers of T cells. What is not known is how expansion affects the ratio of regulatory and effector T cells.

15.5.3.5 T-Cell Stimulation by Whole Therapeutic Proteins

The recognition of whole therapeutic proteins requires the presence of an APC that is capable of processing and presenting peptides derived from the protein [14]. Human monocyte-derived dendritic cells revealed to be very useful with regard to antigen processing *in vitro* closely reflecting antigen processing by professional APCs *in vivo.*

15.5.3.6 T-Cell Responses in Artificial Lymph Nodes

Several artificial lymph node (ALN) systems have been developed to mimic, in three-dimensional structures, the natural lymph node environment of T-cell responses [46,47]. Typically, human blood-derived dendritic cells are cultured in transwells partitioned by human vascular endothelial cells. Addition of autologous $CD4^+$ T cells to the coculture allows the migration of activated APCs through the transwells, mimicking the migration of APCs from the periphery to the lymph nodes. CD154, IFN-γ, IL-2, IL-5, and IL-17 are surrogate markers for activated $CD4^+$ T cells. A good correlation between previously established immune responses *in vivo* and ALN immunogenicity has been observed, at least for protein-based vaccines [48]. Currently attempts are

underway to exploit ALN models for the prediction of immunogenicity of therapeutic proteins.

15.5.4 Mouse Models

To overcome species barriers that limit the translation of results obtained in mice to humans, numerous activities are currently in progress to develop improved murine *in vivo* models. These attempts will be discussed in Sections 15.5.4.1 and 15.5.4.2.

15.5.4.1 HLA Transgenic Mice

The HLA transgenic lines are generated by incorporation of human HLA class II genes into murine MHC class II-deficient mice, generating mouse strains that express selected HLA class II molecules [49,50]. These mice present epitopes in the context of human HLA molecules, and their T cells recognize epitopes presented by HLA class II molecules. HLA transgenic mice are most useful when directly comparing two proteins that are very similar (such as FVIII and versions of FVIII that have fewer epitopes or new glycosylation variants) [36]. A direct correlation has been found between epitopes that elicit T-cell responses in infected humans and those that induce T-cell responses in immunized HLA transgenic mice [51,52].

HLA transgenic mice were successfully used in combination with *in silico* T-cell epitope screening for deimmunizing a therapeutic protein, as reported recently [36]. An *in silico* algorithm localized immunogenic peptides within the C2 domain of the therapeutic protein FVIII. Amino acid changes were introduced within the T-cell epitopes to abolish or at least reduce binding to the human allotype HLA-DR3, as confirmed in a HLA binding assay. Subsequently, HLA-DR3 transgenic mice were used to demonstrate that immunization with the deimmunized epitopes gave rise to significantly reduced immune immunogenicity while the original epitopes were immunogenic [53].

In conclusion, this is a good example of how a combination of *in silico* screening, *in vitro* testing, and HLA transgenic mice can be applied early in development to mitigate the immunogenicity risk of a therapeutic protein that is known to impose an unacceptable immunogenicity risk in its original version.

15.5.4.2 Humanized Mouse Models

Mice engrafted with a functional human immune system, denoted as "humanized," are currently explored to study, among other aspects, human hematopoiesis, stem cell function, and immunogenicity. Immunocompromised SCID/NOD/γ chain$^{-/-}$ or RAG2$^{-/-}$/γ chain$^{-/-}$ mice, utilized as recipients to facilitate acceptance of human tissue, are engrafted with functional human CD34^{+} hematopoietic stem cells, human liver, and human thymus [54]. The result is a cohort of mice in which human myeloid and lymphoid lineages are reconstituted from a single human donor, and the interactions of these cells in a complex biological

environment can be studied. XenoMouse® and the humanized mice such as NOD/Shi-*scid*/IL-2Rγ^{null} mice (NOG), NOD/scid/IL-2 receptor gamma chain knockout mice (NSG), bone marrow, liver, and thymus mice (BLT), and bone marrow transplanted mice (BMT) have all been used as animal models to evaluate human immune responses [54–57].

While humanized mice provide functional and testable elements of the innate and adaptive human immune system without putting patients at risk [58], certain aspects of a fully functional human immune system critical for immunogenicity risk assessment are lacking: the ability to elicit the complete spectrum of B-cell antibody responses or the ability to proteolytically process antigens in the way human endosomal/lysosomal proteases do. Likewise, the humoral immune response in the XenoMouse is restricted by murine MHC and T-cell help but is not as robust as in wild-type mice, potentially due to inefficient signal transduction and isotype switching mediated by accessory factors that are necessary for B-cell maturation. Hence, the utility of such a model to study immune responses to human proteins remains somewhat limited.

In conclusion, more work is required to leverage mouse models to accurately reflect human immune responses before they can become routine tools to assess the immunogenic potential of therapeutic proteins.

15.5.5 Case Studies

15.5.5.1 Translation of *In Silico* and *In Vitro* Data into Clinical Context

Koren *et al.* demonstrated a correlation between the *in silico* evaluation of T helper epitope content of a protein therapeutic and its observed immunogenicity in a clinical trial [33]. The therapeutic protein was a recombinant Fc fusion protein (FPX) consisting of human germline Fcγ fragment with two identical, biologically active, 24-amino-acid peptides attached to the amino-terminal end of the Fc fragment. Clinically, a single subcutaneous or intravenous administration of the fusion protein resulted in a robust ATA response: high-affinity binding antibodies were found in 40 and 33% of total individuals, respectively.

Based on a retrospective *in silico* analysis, the carboxy-terminal region of the peptide scores high for binding to five of eight common HLA molecules, suggesting that this peptide carries considerable immunogenic potential. In parallel, T-cell-mediated recall responses to the therapeutic protein were assessed *in vitro*: PBMC activation by the FPX peptide and the amino-terminal and carboxy-terminal fragments thereof was measured as a function of the number of IFN-γ and IL-4 spot-forming cells (SFCs) in a standard ELISpot assay. The cytokine data supported the *in silico* data and correlated with the ATA data obtained in a clinical trial.

Immune responses both *in vivo* and *in vitro* to different regions of the protein and to different HLA-DR allotypes also support the *in silico* predictions: the carboxy-terminal region of the FPX peptide showed the highest MHC binding score in the context of the HLA-DRB1*0701 allele. T-cell and antibody responses to

this fragment were observed *in vitro* and *in vivo*, respectively, for individuals sharing the HLA-DRB1*0701 allele. In contrast, the DRB1*0301 allele had very low MHC binding scores, and patients who expressed the DRB1*0301 molecule but not any of the other high-binding alleles demonstrated low responses in ELISpot and no evidence of an ATA response. The immunogenicity of the FPX fragments and the association between clinical results and the HLA class II alleles were supported in the naïve blood T-cell assay in further studies performed by Jawa and coworkers [14,33].

In summary, the *in silico* prediction of immunogenic T helper cell epitope(s) within the carboxy-terminal region of the FPX peptide and the data obtained *in vitro* with T-cell activation assays correlated in qualitative and quantitative terms with ATA responses observed in clinical trials.

15.5.5.2 Link between HLA Haplotype and Immunogenicity: *In Vivo* versus *In Vitro*

An association between the HLA-DRB1*0701 allele and a strong ATA response to recombinant IFN-β was reported in a cohort of multiple sclerosis patients [59]. Potential IFN-β T-cell epitopes were identified using a peptide library and a HLA binding assay, involving B-cell lines expressing the relevant HLA-DR alleles. Peptides were synthesized as overlapping 17-mers covering the entire sequence of IFN-β and grouped into 10 peptide pools. Recall responses from patients subjected to IFN-β therapy were assessed *in vitro* by utilizing an IFN-γ ELISpot T-cell assay. T cells from ATA-positive subjects (HLA-DRB1*0701/HLA-DQA1*0201) secreted high levels of IFN-γ in the presence of 2 of the 10 peptide pools tested, while antibody-negative subjects showed no response. Peptide pools could be deconvoluted to identify the minimal T-cell epitope recombinant IFN-β and HLA-DRB1*0701 to be the restricting HLA class II allele.

Stickler *et al.* have observed that the DRB1*1501/DQB1*0602 haplotype is associated with a high potential of naïve $CD4^+$ T cells to react against human IFN-β *in vitro* [60]. In accordance, the C-terminal IFN-β peptide 147–161 was eluted from HLA-DR molecules after IFN-β exposure of human dendritic cells expressing the DRB1*1501 allele (H. Kropshofer, unpublished). Due to limitations in the number of subjects tested, a clinical correlation with the HLA-DRB1*1501 allele could not be detected in the previous study [59].

15.6 Immunogenicity Risk Mitigation

Several different preclinical approaches to mitigate the immunogenicity of therapeutic proteins are currently under consideration: (i) modification of the therapeutic protein by pegylation and/or glycosylation to mask potential T-cell epitopes, thereby reducing recognition by the immune system; (ii) modification of immunodominant T-cell epitopes to remove anchor residues critical for HLA binding, denoted as "deimmunization," thereby disrupting their potential to activate T cells; and (iii) strategies to tolerize the immune system to the therapeutic

protein. The concepts of deimmunization and tolerization will be discussed in more detail. Clinically, it is often feasible to balance aspects that are known to impact the immunogenicity risk, such as the purity of a drug product or the mode of administration, as discussed below.

15.6.1 Deimmunization

The first published attempt to deimmunize a protein involved the introduction of alanine substitutions into the protein staphylokinase, leading to a reduction of the potential to both activate T cells *in vitro* and trigger ATA responses *in vivo* [61]. This pioneering approach was followed by a number of further deimmunization studies [62–65]. Ongoing efforts include deimmunization of botulinum neurotoxin type A, lysostaphin, and factor VIII [36].

It is obvious that the utility of this approach depends on the location and extent of amino acid changes and the impact of those changes on the pharmacological activity of the modified protein. A compromise between the number of residues necessary to abolish HLA binding while maintaining the pharmacological activity of a therapeutic protein may eventually lead to less immunogenic proteins while preserving their therapeutic potential. Several deimmunized therapeutic proteins are currently in clinical evaluation [61]; hence, future clinical trial results will reveal whether deimmunized biotherapeutic proteins display acceptably low incidences of immunogenicity and associated safety events.

15.6.2 Tolerization

Active interference with $CD4^+$ T-cell responses to protein therapeutics by inducing tolerance to the drug is an approach that has attracted significant interest in the recent past. Tolerance induction may be achieved by the use of nondepleting anti-CD4 antibodies [66,67]. An alternative approach employed IVIGs (intravenous immunoglobulins): IVIGs induce tolerance in solid organ transplantation [68], in the context of neutralizing ATAs against FVIII [69], and to inhibit ATA formation in Pompe patients undergoing Myozyme treatment [70]. De Groot *et al.* have identified a set of natural, human regulatory T-cell epitopes ("Tregitopes") present in the Fc and Fab domains of IgG that have also been shown to induce tolerance to coadministered proteins [62]. When incubated with PBMCs *in vitro*, Tregitopes activate $CD4^+CD25^+Foxp3^+$ T cells and increase expression of regulatory cytokines and chemokines [62]. Methods for coadministering Tregitopes with protein therapeutics are currently under development. However, as yet, it is open whether Tregitopes are suitable to induce tolerance against therapeutic proteins in humans when Tregitopes are utilized outside their intrinsic IgG sequence context.

15.6.3
Clinical Control of Immunogenicity Risk Factors

Risk minimization at the preclinical stage is of key importance since only then it is feasible to implement deimmunization and tolerization approaches, as described above. However, also at the clinical development stage it is important to set up risk mitigation strategies prior to running proof-of-concept studies.

In most instances, the indication, the patient population, and the PK/PD of the drug predetermine a number of immunogenicity risk factors, while other risk aspects are still variable, depending on the exact design of a clinical study. For example, in type 2 diabetes (T2D), the genotype and the pre-/comedication of patients are expected not to impose an elevated risk for immunogenicity, nor is this the case from the perspective of typical T2D targets that are unrelated to the immune system. As a consequence, even a combination of subcutaneous administration of a protein therapeutic and intermittent dosing at moderate dose levels may be uncritical with regard to the immunogenicity risk. A successful risk mitigation strategy, however, should not accept additional risk factors, for example, formulation-related risk factors such as particular excipients, an elevated likelihood of formation of aggregates *in situ*, or elevated levels of impurities, such as host cell proteins or manufacturing side products.

In conclusion, immunogenicity risk mitigation starts with the amino acid sequence and the final structure of a therapeutic protein, but there is a need to continue with risk mitigation efforts during early- and late-stage clinical development by balancing risk factors related to the design of the formulation, the manufacturing process, and the clinical studies *per se*.

15.7
The Integrated Strategy of Risk Minimization

Given the potentially severe impact of unwanted immunogenicity on the development of therapeutic protein drugs, there is increasing interest at both the preclinical and clinical stages to have in place an integrated strategy of immunogenicity risk mitigation. A multitiered approach promises most success and is outlined here (Figure 15.3).

Tier 1: HLA association (during lead selection)

At the lead selection stage, the drug candidates shall be subjected to a combination of *in silico* screening and *in vitro* HLA binding or, alternatively, the MAPPs approach. Such a combinatorial approach is mandatory for proteins that are not identical in structure to endogenous proteins, even if only a single amino acid has been changed compared with the native molecule. In case of human or humanized monoclonal antibodies, primarily the CDR regions need to undergo an *in silico* screen or a MAPPs assay [71]. Synthetic peptides reflecting the regions identified as potential T-cell epitopes by the *in silico* test

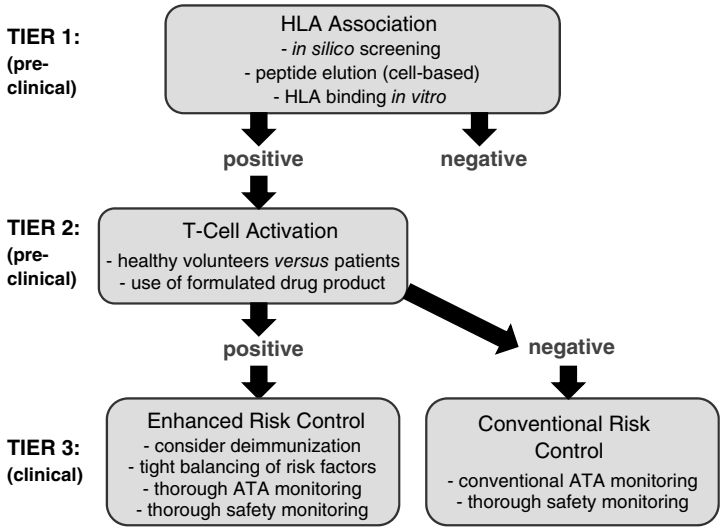


Figure 15.3 The integrated strategy of risk minimization.

need to be synthesized and tested in a HLA binding assay (covering at least 10–15 major HLA haplotypes) to confirm their true nature as T-cell epitopes. In case T-cell epitopes have been identified by the HLA binding assay, these peptides need to be subjected to tier-2 testing. In case no T-cell epitopes were found in tier 1, there is no need for tier-2 testing – tier-3 considerations will apply then (see below).

Tier 2: T-cell activation (during lead selection)

The set of potential T-cell epitopes identified in tier 1 need to be tested for T-cell activation, preferentially in the context of blood cells derived from a representative set of patients rather than healthy volunteers. It is recommended to utilize at least two independent readouts of T-cell activity to ensure a balance between specificity and sensitivity. The added value of the T-cell activation assay is threefold: (i) the use of whole therapeutic protein rather than peptides only may enable the evaluation of processing-associated changes, such as posttranslational modifications, misfolding, or storage-related changes, such as deamidation or oxidation or formation of aggregates; (ii) the use of fully formulated therapeutic protein may allow the detection of formulation-related effects; and (iii) the use of patient-derived T cells and antigen-presenting cells may help to better understand disease-specific characteristics that may impact the immunogenicity risk. In case the therapeutic protein tests positive in tier 2, a deimmunization effort may need to be taken into consideration in order to reduce the immunogenicity risk. If a deimmunization cannot be carried out, because of otherwise negative consequences on the biological activity, more emphasis needs to be put onto tier 3.

Tier 3: Risk control (during clinical development)

The risk mitigation strategy during clinical development needs to be based on the outcome and follow-up activities of the preclinical tier-1 and tier-2 approaches: for example, in case a deimmunization was carried out, the risk is lower compared with the situation when a protein tested positive in tier 1 and tier 2, but a deimmunization could not be done. Likewise, a low ATA incidence ($\ll$10%) in phase I/II indicates a lower risk, while ATA incidences >25% may impose an elevated immunogenicity risk for phase III clinical trials so that the implementation of a risk mitigation plan is more obvious. Depending on which other factors are suggestive of an elevated risk during the development program, for example, presence of a nonredundant endogenous counterpart of the protein drug, target is a structure on immune cells, drug will be dosed lifelong, and/or drug will be used in an autoimmune indication, the risk mitigation plan needs to be designed differently.

15.8 Summary

Since unwanted immunogenicity of therapeutic proteins may trigger serious adverse events – though rare in incidence – FDA and the EMA are recommending that each development program be accompanied by an immunogenicity risk assessment strategy [19]. The recent EMA guidance mentions "predictive immunogenicity" as an approach sponsors could consider in their preclinical studies [19]. The historical focus has been on measurement of ATA responses as the readout for immunogenicity, supported by the obvious consequences of ATA responses on the pharmacokinetics, efficacy, and safety of therapeutic proteins.

At present, drug developers are applying strategies to assess and modulate humoral and/or cell-mediated immune responses directed against protein therapeutics at both the preclinical and clinical phases of development. Guidelines for standardizing immunogenicity testing of protein therapeutics across the industry are emerging from these activities. Efforts to reduce the immunogenic potential of protein drugs are also becoming more and more common – drugs that have been deimmunized or tolerized are expected to enter the clinic within the next few years.

A wide range of immunogenicity testing tools are available for determining whether or not a protein therapeutic is more or less likely to elicit a detrimental immune response in patients. Given the complexity of the immune system, combinations of multiple methods need to be employed to most closely estimate and mitigate immunogenicity risk. As yet, we have leveraged considerably the possibility to localize immunogenicity hotspots within proteins. The next challenge will be to deepen our understanding on what factors link ATA incidence to safety-related events, such as hypersensitivity or infusion reactions. Further evolving tests and the elucidation of safety biomarkers will eventually impact drug development by reducing drug failure, costs to pharmaceutical industry, and risks to patients.

References

1 Barbosa, M.D. (2011) Immunogenicity of biotherapeutics in the context of developing biosimilars and biobetters. *Drug Discov. Today*, **16**, 345–353.

2 Walsh, G. (2010) Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.*, **28**, 917–924.

3 Casadevall, N., Nataf, J., Viron, B., Kolta, A., Kiladjian, J.J., Martin-Dupont, P., Michaud, P., Papo, T., Ugo, V., Teyssandier, I., Varet, B., and Mayeux, P. (2002) Pure red-cell aplasia and anti-erythropoietin antibodies in patients treated with recombinant erythropoietin. *N. Engl. J. Med.*, **346**, 469–475.

4 Li, J., Yang, C., Xia, Y., Bertino, A., Glaspy, J., Roberts, M., and Kuter, D.J. (2001) Thrombocytopenia caused by the development of antibodies to thrombopoietin. *Blood*, **98**, 3241–3248.

5 Chung, C.H., Mirakhur, B., Chan, E., Le, Q.T., Berlin, J., Morse, M., Murphy, B.A., Satinover, S.M., Hosen, J., Mauro, D., Slebos, R.J., Zhou, Q., Gold, D., Hatley, T., Hicklin, D.J., and Platts-Mills, T.A. (2008) Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *N. Engl. J. Med.*, **358**, 1109–1117.

6 Hansel, T.T., Kropshofer, H., Singer, T., Mitchell, J.A., and George, A.J.T. (2010) The safety and side effects of monoclonal antibodies. *Nat. Rev. Drug Discov.*, **10**, 325–338.

7 Buttel, I.C., Chamberlain, P., Chowers, Y., Ehmann, F., Greinacher, A., Jefferis, R., Kramer, D., Kropshofer, H., Lloyd, P., Lubiniecki, A., Krause, R., Mire-Sluis, A., Platts-Mills, T., Ragheb, J.A., Reipert, B.M., Schellekens, H., Seitz, R., Stas, P., Subramanyam, M., Thorpe, R., Trouvin, J. H., Weise, M., Windisch, J., and Schneider, C.K. (2011) Taking immunogenicity assessment of therapeutic proteins to the next level. *Biologicals*, **39**, 100–109.

8 Mire-Sluis, A.R., Barrett, Y.C., Devanarayan, V., Koren, E., Liu, H., Maia, M., Parish, T., Scott, G., Shankar, G., Shores, E., Swanson, S.J., Taniguchi, G., Wierda, D., and Zuckerman, L.A. (2004) Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Methods*, **289**, 1–16.

9 Gupta, S., Indelicato, S.R., Jethwa, V., Kawabata, T., Kelley, M., Mire-Sluis, A.R., Richards, S.M., Rup, B., Shores, E., Swanson, S.J., and Wakshull, E. (2007) Recommendations for the design, optimization, and qualification of cell-based assays used for the detection of neutralizing antibody responses elicited to biological therapeutics. *J. Immunol. Methods*, **321**, 1–18.

10 Shankar, G., Devanarayan, V., Amaravadi, L., Barrett, Y.C., Bowsher, R., Finco-Kent, D., Fiscella, M., Gorovits, B., Kirschner, S., Moxness, M., Parish, T., Quarmby, V., Smith, H., Smith, W., Zuckerman, L.A., and Koren, E. (2008) Recommendations for the validation of immunoassays used for detection of host antibodies against biotechnology products. *J. Pharm. Biomed. Anal.*, **48**, 1267–1281.

11 Stickler, M., Rochanayon, N., Razo, O.J., Mucha, J., Gebel, W., Faravashi, N., Chin, R., Holmes, S., and Harding, F.A. (2004) An *in vitro* human cell-based assay to rank the relative immunogenicity of proteins. *Toxicol. Sci.*, **77**, 280–289.

12 Perry, L.C., Jones, T.D., and Baker, M.P. (2008) New approaches to prediction of immune responses to therapeutic proteins during preclinical development. *Drugs R&D*, **9**, 385–396.

13 Stas, P. and Lasters, I. (2009) Strategies for preclinical immunogenicity assessment of protein therapeutics. *IDrugs*, **12**, 169–173.

14 Wullner, D., Zhou, L., Bramhall, E., Kuck, A., Goletz, T.J., Swanson, S., Chirmule, N., and Jawa, V. (2010) Considerations for optimization and validation of an *in vitro* PBMC derived T cell assay for immunogenicity prediction of biotherapeutics. *Clin. Immunol.*, **137**, 5–14.

15 EMA (2007) Guideline on Immunogenicity Assessment of Biotechnology-Derived Therapeutic Proteins. EMEA/CHMP/BMWP/14327/2007.

16 Mueller, R., Karle, A., Vogt, A., Kropshofer, H., Ross, A., Maeder, K., and Mahler, H.C. (2009) Evaluation of the immuno-stimulatory potential of stopper extractables and leachables by using dendritic cells as readout. *J. Pharm. Sci.*, **98**, 3548–3561.
17 Rosenberg, A.S. (2006) Effects of protein aggregates: an immunologic perspective. *AAPS J.*, **8**, E501–E507.
18 Verthelyi, D. and Wang, V. (2010) Trace levels of innate immune response modulating impurities (IIRMIs) synergize to break tolerance to therapeutic proteins. *PLoS One*, **5**, e15252.
19 Singh, S.K. (2011) Impact of product-related factors on immunogenicity of biotherapeutics. *J. Pharm. Sci.*, **100**, 354–387.
20 Rathore, A.S. and Winkler, H. (2007) Quality by design for biopharmaceuticals. *Nat. Biotechnol.*, **27**, 26–34.
21 Zubler, R.H. (2001) Naive and memory B cells in T-cell-dependent and T-independent responses. *Springer Semin. Immunopathol.*, **23**, 405–419.
22 Vogt, A.B., Ploix, C., and Kropshofer, H. (2005) Antigen processing for MHC class II molecules, in *Antigen Presenting Cells* (eds H. Kropshofer and A.B. Vogt), Wiley-VCH Verlag GmbH, Weinheim, pp. 89–116.
23 Pratt, K.P. and Thompson, A.R. (2009) B-cell and T-cell epitopes in anti-factor VIII immune responses. *Clin. Rev. Allergy Immunol.*, **37**, 80–95.
24 Kropshofer, H., Vogt, A.B., Moldenhauer, G., Hammer, J., Blum, J.S., and Hammerling, G.J. (1996) Editing of the HLA-DR peptide repertoire by HLA-DM. *EMBO J.*, **15**, 6144–6154.
25 Eyerman, M.C., Zhang, X., and Wysocki, L.J. (1996) T cell recognition and tolerance of antibody diversity. *J. Immunol.*, **157**, 1037–1046.
26 Bluestone, J.A. and Abbas, A.K. (2003) Natural versus adaptive regulatory T cells. *Nat. Rev. Immunol.*, **3**, 253–257.
27 Tsuji, N.M. and Kosaka, A. (2008) Oral tolerance: intestinal homeostasis and antigen-specific regulatory T cells. *Trends Immunol.*, **29**, 532–540.
28 Congia, M., Patel, S., Cope, A.P., De Virgilis, S., and Sonderstrup, G. (1998) T cell epitopes of insulin defined in HLA-DR4 transgenic mice are derived from preproinsulin and proinsulin. *Proc. Natl. Acad. Sci. USA*, **95**, 3833–3838.
29 Kropshofer, H. and Spindeldreher, S. (2005) in *Antigen Presenting Cells* (eds H. Kropshofer and A.B. Vogt), Wiley-VCH Verlag GmbH, Weinheim, pp. 159–189.
30 Bryson, C.J., Jones, T.D., and Baker, M.P. (2010) Prediction of immunogenicity of therapeutic proteins: validity of computational tools. *BioDrugs*, **24**, 1–8.
31 De Groot, A.S., Bosma, A., Chinai, N., Frost, J., Jesdale, B.M., Gonzalez, M.A., Martin, W., and Saint-Aubin, C. (2001) From genome to vaccine: *in silico* predictions, *ex vivo* verification. *Vaccine*, **19**, 4385–4395.
32 Inaba, H., Martin, W., De Groot, A.S., Qin, S., and De Groot, L.J. (2006) Thyrotropin receptor epitopes and their relation to histocompatibility leukocyte antigen-DR molecules in Graves' disease. *J. Clin. Endocrinol. Metab.*, **91**, 2286–2294.
33 Koren, E., De Groot, A.S., Jawa, V., Beck, K.D., Boone, T., Rivera, D., Li, L., Mytych, D., Koscec, M., Weeraratne, D., Swanson, S., and Martin, W. (2007) Clinical validation of the "*in silico*" prediction of immunogenicity of a human recombinant therapeutic protein. *Clin. Immunol.*, **124**, 26–32.
34 Doytchinova, I.A., Walshe, V.A., Jones, N. A., Gloster, S.E., Borrow, P., and Flower, D.R. (2004) Coupling *in silico* and *in vitro* analysis of peptide-MHC binding: a bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes. *J. Immunol.*, **172**, 7495–7502.
35 De Groot, A.S. and Martin, W. (2009) Reducing risk, improving outcomes: bioengineering less immunogenic protein therapeutics. *Clin. Immunol.*, **131**, 189–201.
36 Moise, L., Song, C., Martin, W.D., Tassone, R., De Groot, A.S., and Scott, D.W. (2012) Effect of HLA DR epitope de-immunization of Factor VIII *in vitro* and *in vivo*. *Clin. Immunol.*, **142**, 320–331.

37 Kropshofer, H. and Singer, T. (2006) Overview of cell-based tools for pre-clinical assessment of immunogenicity of biotherapeutics. *J. Immunotoxicol.*, **3**, 131–136.

38 Rohn, T.A., Reitz, A., Paschen, A., Nguyen, X.D., Schadendorf, D., and Vogt, A.B., and Kropshofer, H. (2005) A novel strategy for the discovery of MHC class II-restricted tumor antigens: identification of a melanotransferrin helper T-cell epitope. *Cancer Res.*, **65**, 10068–10078.

39 Kessler, J.H., Mommaas, B., Mutis, T., Huijbers, I., Vissers, D., Benckhuijsen, W.E., Schreuder, G.M., Offringa, R., Goulmy, E., Melief, C.J., van der Burg, S.H., and Drijfhout, J.W. (2003) Competition-based cellular peptide binding assays for 13 prevalent HLA class I alleles using fluorescein-labeled synthetic peptides. *Hum. Immunol.*, **64**, 245–255.

40 Buchli, R., Van Gundy, R.S., Hickman-Miller, H.D., Giberson, C.F., Bardet, W., and Hildebrand, W.H. (2005) Development and validation of a fluorescence polarization-based competitive peptide-binding assay for HLA-A*0201 – a new tool for epitope discovery. *Biochemistry*, **44**, 12491–12507.

41 Wu, L.C., Tuot, D.S., Lyons, D.S., Garcia, K.C., and Davis, M.M. (2002) Two-step binding mechanism for T-cell receptor recognition of peptide MHC. *Nature*, **418**, 552–556.

42 Makitalo, B., Andersson, M., Arestrom, I., Karlen, K., Villinger, F., Ansari, A., Paulie, S., Thorstensson, R., and Ahlborg, N. (2002) ELISpot and ELISA analysis of spontaneous, mitogen-induced and antigen-specific cytokine production in cynomolgus and rhesus macaques. *J. Immunol. Methods*, **270**, 85–97.

43 Carson, R.T. and Vignali, D.A. (1999) Simultaneous quantitation of 15 cytokines using a multiplexed flow cytometric assay. *J. Immunol. Methods*, **227**, 41–52.

44 Stickler, M.M., Estell, D.A., and Harding, F.A. (2000) $CD4^+$ T-cell epitope determination using unexposed human donor peripheral blood mononuclear cells. *J. Immunother.*, **23**, 654–660.

45 Delluc, S., Ravot, G., and Maillere, B. (2010) Quantification of the preexisting CD4 T-cell repertoire specific for human erythropoietin reveals its immunogenicity potential. *Blood*, **116**, 4542–4545.

46 Giese, C., Lubitz, A., Demmler, C.D., Reuschel, J., Bergner, K., and Marx, U. (2010) Immunological substance testing on human lymphatic micro-organoids *in vitro*. *J. Biotechnol.*, **148**, 38–45.

47 Ma, Y., Poisson, L., Sanchez-Schmitz, G., Pawar, S., Qu, C., Randolph, G.J., Warren, W.L., Mishkin, E.M., and Higbee, R.G. (2010) Assessing the immunopotency of Toll-like receptor agonists in an *in vitro* tissue-engineered immunological model. *Immunology*, **130**, 374–387.

48 Schanen, B.C., De Groot, A.S., Moise, L., Ardito, M., McClaine, E., Martin, W., Wittman, V., Warren, W.L., and Drake, D.R. (2011) Coupling sensitive *in vitro* and *in silico* techniques to assess cross-reactive $CD4^+$ T cells against the swine-origin H1N1 influenza virus. *Vaccine*, **29**, 3299–3309.

49 Pan, S., Trejo, T., Hansen, J., Smart, M., and David, C.S. (1998) HLA-DR4 (DRB1*0401) transgenic mice expressing an altered CD4-binding site: specificity and magnitude of DR4-restricted T cell response. *J. Immunol.*, **161**, 2925–2929.

50 Kong, Y.C., Lomo, L.C., Motte, R.W., Giraldo, A.A., Baisch, J., Strauss, G., Hammerling, G.J., and David, C.S. (1996) HLA-DRB1 polymorphism determines susceptibility to autoimmune thyroiditis in transgenic mice: definitive association with HLA-DRB1*0301 (DR3) gene. *J. Exp. Med.*, **184**, 1167–1172.

51 Shirai, M., Arichi, T., Nishioka, M., Nomura, T., Ikeda, K., Kawanishi, K., Engelhard, V.H., Feinstone, S.M., and Berzofsky, J.A. (1995) CTL responses of HLA-A2.1-transgenic mice specific for hepatitis C viral peptides predict epitopes for CTL of humans carrying HLA-A2.1. *J. Immunol.*, **154**, 2733–2742.

52 Le, A.XRRCR.R., Bernhard, E.J., Holterman, M.J., Strub, S., Parham, P., Lacy, E., and Engelhard, V.H. (1989) Cytotoxic T cell responses in HLA-A2.1 transgenic mice. Recognition of HLA

alloantigens and utilization of HLA-A2.1 as a restriction element. *J. Immunol.*, **142**, 1366–1371.

53 Mangalam, A., Rodriguez, M., and David, C. (2007) A new humanized HLA transgenic mouse model of multiple sclerosis expressing class II on mouse CD4T cells. *Ann. N. Y. Acad. Sci.*, **1103**, 112–117.

54 Brehm, M.A., Cuthbert, A., Yang, C., Miller, D.M., DiIorio, P., Laning, J., Burzenski, L., Gott, B., Foreman, O., Kavirayani, A., Herlihy, M., Rossini, A.A., Shultz, L.D., and Greiner, D.L. (2010) Parameters for establishing humanized mouse models to study human immunity: analysis of human hematopoietic stem cell engraftment in three immunodeficient strains of mice bearing the IL2rgamma (null) mutation. *Clin. Immunol.*, **135**, 84–98.

55 Ishikawa, F., Saito, Y., Yoshida, S., Harada, M., and Shultz, L.D. (2008) The differentiative and regenerative properties of human hematopoietic stem/progenitor cells in NOD-SCID/IL2rgamma(null) mice. *Curr. Top. Microbiol. Immunol.*, **324**, 87–94.

56 Wege, A.K., Melkus, M.W., Denton, P.W., Estes, J.D., and Garcia, J.V. (2008) Functional and phenotypic characterization of the humanized BLT mouse model. *Curr. Top. Microbiol. Immunol.*, **324**, 149–165.

57 Reddy, P., Negrin, R., and Hill, G.R. (2008) Mouse models of bone marrow transplantation. *Biol. Blood Marrow Transplant.*, **14**, 129–135.

58 Willinger, T., Rongvaux, A., Strowig, T., Manz, M.G., and Flavell, R.A. (2011) Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement. *Trends Immunol.*, **32**, 321–327.

59 Barbosa, M.D., Vielmetter, J., Chu, S., Smith, D.D., and Jacinto, J. (2006) Clinical link between MHC class II haplotype and interferon-beta (IFN-beta) immunogenicity. *Clin. Immunol.*, **118**, 42–50.

60 Stickler, M., Valdes, A.M., Gebel, W., Razo, O.J., Faravashi, N., Chin, R., Rochanayon, N., and Harding, F.A. (2004) The HLA-DR2 haplotype is associated with an increased proliferative response to the immunodominant $CD4^+$ T-cell epitope in human interferon-beta. *Genes Immun.*, **5**, 1–7.

61 Warmerdam, P.A., Plaisance, S., Vanderlick, K., Vandervoort, P., Brepoels, K., Collen, D., and De Maeyer, M. (2002) Elimination of a human T-cell region in staphylokinase by T-cell screening and computer modeling. *Thromb. Haemost.*, **87**, 666–673.

62 De Groot, A.S., Moise, L., McMurry, J.A., Wambre, E., Van Overtvelt, L., Moingeon, P., Scott, D.W., and Martin, W. (2008) Activation of natural regulatory T cells by IgG Fc-derived peptide "Tregitopes". *Blood*, **112**, 3303–3311.

63 Parker, A.S., Zheng, W., Griswold, K.E., and Bailey-Kellogg, C. (2010) Optimization algorithms for functional deimmunization of therapeutic proteins. *BMC Bioinformatics*, **11**, 180.

64 Cantor, J.R., Yoo, T.H., Dixit, A., Iverson, B.L., Forsthuber, T.G., and Georgiou, G. (2011) Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift. *Proc. Natl. Acad. Sci. USA*, **108**, 1272–1277.

65 De Groot, A.S., Knopp, P.M., and Martin, W. (2005) De-immunization of therapeutic proteins by T-cell epitope modification. *Dev. Biol. (Basel)*, **122**, 171–194.

66 Cobbold, S.P., Qin, S.X., and Waldmann, H. (1990) Reprogramming the immune system for tolerance with monoclonal antibodies. *Semin. Immunol.*, **2**, 377–387.

67 Winsor-Hines, D., Merrill, C., O'Mahony, M., Rao, P.E., Cobbold, S.P., Waldmann, H., Ringler, D.J., and Ponath, P.D. (2004) Induction of immunological tolerance/hyporesponsiveness in baboons with a nondepleting CD4 antibody. *J. Immunol.*, **173**, 4715–4723.

68 Jordan, S.C., Toyoda, M., and Vo, A.A. (2011) Regulation of immunity and inflammation by intravenous immunoglobulin: relevance to solid

organ transplantation. *Expert Rev. Clin. Immunol.*, **7**, 341–348.

69 Zandman-Goddard, G., Krauthammer, A., Levy, Y., Langevitz, P., and Shoenfeld, Y. (2011) Long-term therapy with intravenous immunoglobulin is beneficial in patients with autoimmune diseases. *Clin. Rev. Allergy Immunol.*, **52**, 247–255.

70 Mendelsohn, N.J., Messinger, Y.H., Rosenberg, A.S., and Kishnani, P.S. (2009) Elimination of antibodies to recombinant enzyme in Pompe's disease. *N. Engl. J. Med.*, **360**, 194–195.

71 Harding, F.A., Stickler, M.M., Razo, J., and DuBridge, R.B. (2010) The immunogenicity of humanized and fully human antibodies: residual immunogenicity resides in the CDR regions. *mAbs*, **2**, 256–265.

16
Regulatory Aspects

Beatriz Silva Lima

16.1
The History of Medicines Regulations in Brief

16.1.1
United States of America

The history of drug regulation in the United Sates and Europe has been shaped by accidents with medicinal products. The sulfanilamide disaster was decisive for the amendment of the US Food and Drug Act dated 1906. In 1937, around 107 deaths, mostly children (the target population), occurred following the use of a sulfanilamide liquid formulation containing polyethylene glycol as solvent. The new Federal Food, Drug, and Cosmetic (FDC) Act of 1938 contained a new request for proof of safety of new drugs before reaching the market [1]. Subsequently, the thalidomide disaster, which occurred in Europe in the late 1950s, had a major impact on the US drug regulations. Thalidomide was synthesized in 1954. In 1958, it headed the over-the-counter sales in West Germany as new sedative pills, and in 1961 it was marketed in 46 countries throughout Europe, also used to reduce nausea in pregnant women. The drug-induced birth defects occurred in about 13 000 babies, mostly in Germany and England. Babies were without hands, feet, toes, or fingers, like flippers growing out of their shoulders and trunk (phocomelia). Although not marketed in the United States (it was refused by FDA, thanks to Dr. Frances Kelsey), it had been distributed out of market, seeming that 40 malformed babies were born [2]. Triggered by thalidomide, in 1962 the Kefauver Harris amendment quoted that animal models be used to predict human response and avoid another tragedy [3]. Thalidomide teratogenicity could not be reproduced in the rat and the mouse, and was only observed in the rabbit, although not in all strains [3]. Consequently, the rabbit (or a nonrodent species) has been included as one of the two species for teratogenicity studies of all new drugs. This request still persists, at least for new small molecules. Interestingly, while triggering this quote, it is worth referring that animal predictivity for thalidomide teratogenicity is poor, even the White New Zealand rabbit demonstrated phocomelia only at a dose between 25 and

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

Table 16.1 Preclinical tests required by the Division in "Procedures for the appraisal of the toxicity of chemicals in foods," September 1949 [5].

A. Pharmacodynamics	Including measurements of blood pressure, respiration, and heart rate. The results of these studies could be used to plan further tests
B. Acute toxicity	The determination of this index involved compiling a dose–response curve, using a minimum of three species for comparison of symptoms
C. Subacute toxicity	Daily doses to be given to one or more species for 6–12 weeks to be used as a guide in the design of chronic experiments
D. Chronic tests	Chronic tests, three or more species to be used, one for the duration of its lifetime (rats suggested)
E. External effects	External effects, sensitization, skin irritation, and so on
F. Special studies	Special studies such as reproduction, hematology, absorption, and excretion

300 times that given to humans [3], and in multiple studies and species, thalidomide has revealed teratogenicity only occasionally [4]. Following the Second World War, the FDA recommended an extensive battery of tests for chemicals relying on animal toxicity studies, justified by the inevitability of exposure to novel synthetic chemicals (Table 16.1).

Only in 1955 specific guidance pertaining to medicinal products was incorporated in the FDA guidance [6]. Studies requested for marketing authorization of medicines and chemicals were aligned. Later in 1962, carcinogenicity studies had been established and guidelines for toxicity tests for all drugs were written (known as the Lehman Guidelines) to assist the pharmaceutical industry to comply with the new law. The standardized carcinogenicity protocols developed were the basis for the current 2-year rodent carcinogenicity studies [6].

16.1.2 Europe

In European countries, before the thalidomide tragedy, it was not mandatory to submit proofs of safety, efficacy, or quality data to Health Authorities prior to marketing of medicinal products. The thalidomide disaster led to the reshaping of the regulatory system in countries such as Germany (where the manufacturer of thalidomide was based) and the United Kingdom [7]. In the European Union, the harmonization of the requirements for drug approval started in 1965 with the introduction of several Council Directives [8–11]. Through the latter, the Committee for Proprietary Medicinal Products (CPMP) was established as an advisory committee to the European Community. These directives were the starting point for the "common market" for medicines. In 1993, the European Medicines Evaluation Agency (EMEA) was created, with the CPMP, currently CHMP (Committee for Human Medicinal Products), to formulate the Agency opinions on centralized applications and granting marketing authorizations [12]. Harmonized guidelines on nonclinical safety requirements for marketing medicinal products in the European Community were published in 1989 [13].

The main nonclinical studies in the document were in line with those established in the United States, including single- and repeated-dose toxicity, genotoxicity, reproductive toxicity, carcinogenicity, and pharmacokinetic and metabolic studies.

16.1.3 The International Conference on Harmonisation

The harmonization of the requirements for medicines approval achieved in the European Community has evidenced its feasibility and led to the expansion across the major regions of the world. The International Conference on Harmonisation (ICH) started in 1990 aiming at overcoming discrepancies in the requirements for the marketing authorization applications in the United States, Japan, and Europe [14]. The most reasonable, scientifically sound approaches are adopted based on the state of the art and the shared experience among the stakeholders: Pharmaceutical Industry Associations/Regulatory Agencies in the United States (Pharma/FDA), Europe (EFPIA/EMA), and Japan (JPMA/PMDA). Major achievements have been reached through harmonized guidelines, which facilitated the development and marketing of medicinal products in all regions involved, based on the same data. The harmonized preclinical paradigm currently in place is well defined in the ICH M3(R2) guideline, which addresses all the nonclinical safety studies needed to support the clinical trials and marketing authorization of medicinal products. The core study program determined in the ICH M3(R2) guideline involves (i) pharmacodynamic studies to characterize the primary, secondary, and safety pharmacology performed *in vitro* and *in vivo*, including animal models of disease; (ii) pharmacokinetic studies in multiple species; (iii) general toxicity studies (repeated-dose toxicology studies) conducted in rodents and nonrodents, whose durations depend on the duration and type of clinical trials (3–6 months in rodents and up to 9 months in nonrodents); (iv) reproductive toxicity studies; and (v) carcinogenicity studies when applicable (e.g., for drugs to be used more than 3 months in a lifetime). In addition, other studies may be needed addressing, for example, neurotoxicity, immunotoxicity, and phototoxicity, anticipated by specific molecular attributes or from findings observed in the core studies. Despite the advances reached within ICH, few changes were introduced in the design and content of the basic preclinical paradigm, in place in the United States since 1955/1962 and in the European Union since 1983/1989. Indeed, the main changes introduced in the last half-century are mostly adjustments of study protocols, criteria for species selection, introduction of innovative biochemical parameters, or new study requirements, such as immunotoxicity or toxicokinetics, or the exploratory clinical trials. Therefore, the ICH requirements mostly reflect the technological advances incorporating, for example, imaging, analytical tools, omics, and so on, but the mostly animal-based test battery has not substantially changed. One area that is under revision and may undergo substantial changes is the carcinogenicity (ICH S1 topic), based on the experience gained by industry and regulators with the rodent

Table 16.2 Design of exploratory clinical trials according to ICH M3(R2) guideline.

Designation	Trial design
Approach 1	Total dose ≤100 μg (no interdose interval limitations) and total dose ≤1/100th of the NOAEL and ≤1/100th of the pharmacologically active dose (scaled on mg/kg for i.v. and mg/m^2 for oral)
Approach 2	Total cumulative dose ≤500 μg, maximum of five administrations with a washout between doses (six or more actual or predicted half-lives) and each dose ≤100 μg and each dose ≤1/100th of the NOAEL and ≤1/100th of the pharmacologically active dose
Approach 3	Single-dose studies at subtherapeutic doses or in the anticipated therapeutic range
Approach 4	Dosing up to 14 days in the therapeutic range but not intended to evaluate clinical MTD
Approach 5	Dosing up to 14 days and not to exceed duration of dosing in nonrodents; in therapeutic range but not intended to evaluate clinical MTD

carcinogenicity studies, and the mechanisms identified as well as new tools under development (*in vitro*, for example, omics). The revised strategy faces the prediction rather than the observation of tumors in rodents and the human relevance of eventual positives, taking into consideration the human benefits and risks. The carcinogenicity assessment would/will be based on an integrated evaluation of the drug attributes, its structural features, biological activity, mitogenic properties, promoting activity, and so on, through, for example, well-established *in vitro* systems, together with the accumulated knowledge on multiple targets and their associated effects (drug pharmacology). The "weight-of-evidence" approach, which puts together all collected evidence, would/will desirably supersede the lifelong rodent study. Adding to these innovative proposals, omics-based approaches, based on the identification of cancer biomarkers, are the subject of dedicated projects, as is the case for the IMI-sponsored MARCAR project "Biomarkers and molecular tumour classification for non-genotoxic carcinogenesis" [15]. Therefore, the value of conducting lifelong rodent studies to assess carcinogenicity as a default is being challenged, and in many situations it could be waived. Such approach is already accepted for biologics (Section 16.4.1). Also, some innovative approaches for clinical trials, potentially affecting the preclinical paradigm, have been incorporated in the ICH M3(R2) guideline [16]. Those trials are exploratory in nature, and can be used during the screening phases for selecting lead candidates. The main protocols are summarized in Table 16.2.

These new type of proof-of-concept clinical studies constitute an important modernization of the current paradigm, opening the opportunity for including human data in the preclinical phases, supported by adapted and reduced preclinical study packages compared with the "classical" phase I–III clinical studies. They may provide information on pharmacodynamics/pharmacokinetics relevant for drug selection or development, for example, testing safety or efficacy biomarkers. In the context of ICH, safety discussions are regularly organized to

evaluate strategies and to work on new opportunities for guidance adaptations or innovations, in line with technological and scientific advances. However, the hurdle is high. Indeed, any modifications of existing requirements, based on new paradigms, can only be accepted by the Regulatory Authorities if there is proof of at least similar predictive potential for human effects, or other type of advantages in line with the principles of reduction, refinement, or replacement of the existing animal tests (3Rs).

16.2 Impact on Drug Success of the Current ICH Nonclinical Testing Paradigm

The expectations for ICH-based regulatory harmonization to improve drug development and reduce attrition have not transferred to reality. High attrition rates are still observed at late stages of drug development, during phase II or III clinical trials or in the postmarketing setting. In a study where the causes and incidences of reasons for drug failure during the development were identified and discussed, based on data collected between 1991 and 2000 [17], the authors concluded that the probability of a drug candidate passing from preclinical stages (i.e., the first GLP toxicity study) to market is 6% or less. The most common factors resulting in project failure were (i) a lack of efficacy (25%), (ii) clinical safety concerns (12%), and (iii) toxicological findings in preclinical evaluation (20%). A survey covering attrition rates for 2011–2012 in phase II and III studies suggests that the rate for failure at phase II has increased (success rates were below 20%) while that at phase III has decreased, which suggests that studies might be designed to lead to earlier termination, if needed. Global reasons for failure were higher for efficacy than for safety [18]. As extensively described in the previous chapters, from implementation of drug preclinical requirements until now, a relevant number of disciplines such as omics, molecular biology, and technologies such as imaging and computational tools associated with improved analytical methods have emerged, and have been incorporated in the study protocols to improve their accuracy and predictive properties. At a high level, it can be commented that, in the past 50 years, the drug development process became more complex but underwent a few major changes in its general format. However, the study has shown that, in contrast to efficacy and safety, the attrition rate due to pharmacokinetics has clearly decreased in the period analyzed. Indeed, the introduction and use of *in vitro* systems to characterize metabolism, *in vitro* and *in silico* models to simulate, predict, and characterize the absorption and biodisposition, and the creation of the innovative exploratory clinical studies for screening (in ICH M3(R2) guideline) led to improved prediction of human pharmacokinetics. The implementation of these clinical studies illustrates that paradigm adjustments based on modernized innovative approaches are possible and useful. In conclusion, paradigm harmonization through ICH appears highly beneficial to avoid study repetitions and align with state of the art, but is not resulting in a substantial format change, despite its

enrichment with technological innovations and new disciplines. It is therefore not surprising that high attrition rates persist.

16.3 Actions Taken for Increasing the Drug Development Success

To reduce the attrition rates in the pharmaceutical field, identifying the underlying reasons is crucial. The predictive value of animal models for relevant human toxicities (HTs) in clinical trials has been analyzed by Olson *et al.* [19]. A multinational pharmaceutical company survey analyzed the concordance between the HTs of pharmaceuticals during development leading to project termination and those observed in animals during the preclinical development. The highest rate of project termination due to HTs was related to (in order) urogenital, cutaneous, hepatic, and cardiovascular side effects. More than 50% of the projects were terminated due to HTs at a late stage of development, 39% in phase I, 43% in phase II, and 10% in phase III. Furthermore, a high percentage of those HTs had been due to the exaggerated pharmacology, in 35% of the cases in phase I, 39% in phase II, and 43% in phase III. The need for a better understanding of the pharmacological pathways is clear from these data. Concordance between human and animal data was seen in 63% of nonrodent studies (primarily the dog), 43% of rodent studies (primarily the rat), and 70% for one or more preclinical animal model species (either in safety pharmacology or in safety toxicology) showing target organ toxicity in the same organ system as the HT. The less predicted toxicities were liver and cutaneous (hypersensitivity). In Europe, an extensive long-term consultation with stakeholders in the biomedical R&D process commenced in October 2004, organized by the European Commission in Brussels to address the causes of delay or bottlenecks. The R&D bottlenecks have been identified as (i) predicting safety, (ii) predicting efficacy, (iii) bridging gaps in knowledge management, and (iv) bridging gaps in education and training. A Strategic Research Agenda (SRA) has been prepared describing the recommendations to address those bottlenecks and a plan for their implementation [20]. With respect to safety and efficacy, increased basic and clinical knowledge as well as strategic measures were discussed as summarized in Table 16.3.

It has been concluded that, for improving the prediction of drug efficacy and safety, increased basic knowledge on several areas is needed; for example, basic mechanisms of disease and involved targets, target biology and associated cascades, target crosstalk, and cascade interconnection would need to be explored. Furthermore, additional and/or alternative preclinical models beyond animal models would be needed. The advances in knowledge and methodologies for cell culturing, including human cells, stem cells, and induced pluripotent stem cells, were considered as potentially providing new testing systems to explore drug safety components. Simultaneously, the developments in the area of genomics, metabolomics, and proteomics are also offering the possibility to explore predictive tools to be used in animals and/or humans to better anticipate or

Table 16.3 Measures to improve the predictivity of safety evaluation [20].

Nonclinical	Clinical
Framework to develop biomarkers to indicate the human relevance and regulatory utility of early laboratory findings	Optimize data resources and strengthen the evidence base in pharmacovigilance
Study the relevance of rodent nongenotoxic carcinogens	Explore the implications of intractable toxicity in animals for human risk
In silico methods for predicting conventional and recently recognized types of toxicity	Develop and strengthen methodologies and networks for pharmacovigilance
Develop better understanding of disease mechanisms	Develop novel methods of risk prediction and benefit–risk assessment
Develop *in vitro* and *in vivo* models predictive of clinical efficacy	Train and educate health care professionals and patients
Develop *in silico* simulations of disease pathology	Stimulation of translational medicine in an integrated fashion across industry and academia
	Creation of disease-specific European imaging networks (for standards, validation of imaging biomarkers)
	Creation of disease-specific European centers for the validation of omics-based biomarkers
	Developing national patient networks and databases (pan-European organization for patient selection and clinical trial analysis)
	Partnership with regulators

monitor drug efficacy and safety. All these possibilities would drive the involvement of basic scientists from academia to help in setting and exploring new basic science-based approaches to improve drug development success, implying the sharing of precompetitive knowledge by pharmaceutical companies. In line with these thoughts, joint actions emerged putting together scientists from academia and industry working with regulatory scientists from regulatory agencies in exploring these innovative ideas and the Innovative Medicines Initiative (IMI) has been created. IMI consists of a public–private partnership sponsored by the European Commission and the EFPIA, to sponsor research aiming to generate strategies and knowledge toward faster discovery and development of better medicines for patients and enhancing Europe's competitiveness [15]. Impressive outcomes were reached with the first 5-year grants, as basis for innovative strategies for drug development, including modeling and simulation, *in vitro* systems with human cells, search for predictive animal and human safety biomarkers (including rodent carcinogenicity), identification of disease targets for CNS, diabetes, and other diseases, and conceiving new approaches for clinical development in identified areas. The IMI project is planned for additional 10 years (IMI 2), in expanded areas under discussion [15]. In the United States, in 2006 the PSTC (Predictive Safety Testing Consortium) has been formed, bringing together pharmaceutical companies to share and validate innovative safety

testing methods under the advisement of the FDA, EMA, PMDA (Japanese Pharmaceutical and Medical Devices Agency), and more than 250 participating scientists (industry and academia). The mission of PSTC is to identify new and improved safety testing methods and submit them for formal regulatory qualification by the FDA, EMA, and PMDA. Preclinical and clinical safety biomarker programs include six working groups, cardiac hypertrophy, kidney, liver, skeletal muscle, testicular toxicity, and vascular injury, and have a strong translational focus to select new safety tools applicable across the drug development spectrum [21]. FDA and EMA qualification has been granted for several urinary protein renal safety biomarkers, for their use in preclinical rodent studies as markers of specific sites for renal toxicity [22,23]. The qualification implies the regulatory assessment and acceptance of the biomarker for the claimed purposes. In total, the output from those and many other initiatives is outstanding and ultimately is expected to lead to the modernization of the development paradigm, with a shift toward more human-relevant approaches.

16.4 Innovative Drugs: Impact on Nonclinical Development Strategies

16.4.1 Biopharmaceuticals

In the last two decades, biopharmaceuticals emerged as innovative therapeutics based on human-specific proteins, for example, insulin, or the human-specific, target-specific monoclonal antibodies (mAbs). The human specificity of those molecules has created difficulties in the selection of relevant animal models to fulfill the existing preclinical paradigm, due to poor activity or differences in the target function or involvement in the intended disease. Rodents and dogs are often not suitable for toxicity testing of biologics, and many assessments are performed in the nonhuman primates (NHPs). Also, since many of those products are immunogenic in animals, the duration of toxicity studies became limited by the emergence of anti-drug antibodies and long-term toxicity studies, even in the NHPs, as well as carcinogenicity studies were unfeasible in many situations. Paradigm adaptations had therefore to be processed, and a specific ICH topic addressing the preclinical development of biotechnology-derived medicinal products was created (ICH S6) in 1995 [24]. Since then, the experience gained so far has shown that mostly the safety concerns are associated with immunogenicity or with excessive pharmacology, and can be predicted, making animal testing less necessary [25]. Carcinogenicity studies are often not feasible, but based on the understanding of the mode of action (MoA) and target involvement, it has been recognized that carcinogenicity can still be predicted mostly taking into consideration the general knowledge and the pharmacology of the molecules. The inadequacy of the classical paradigm to support the estimation of the starting dose in first-in-human (FIH) clinical trials with biopharmaceuticals

became apparent with the mAb TGN1412 incident in 2006. Following the administration of the first human dose calculated based on the NOAEL from a NHP toxicity study, a massive cytokine release (storm) occurred in all the six subjects in the FIH trial. Later, it has been recognized that full receptor occupancy had been reached with the FIH dose used (calculated using the mAb–target binding parameters). The need to reformulate the criteria for judging species relevance as well as the usefulness of pharmacodynamics for human safety estimation became clear with this incident. A new guideline by the EMA on strategies to mitigate the risk in first-in-human clinical trials was published [26]. Furthermore, based on experience gained since 1995 with biopharmaceuticals, updates were introduced in the preclinical requirements in general leading to an addendum to the ICH guideline (ICH S6(R1)) [24]. For instance, for carcinogenicity assessment, alternatives to *in vivo* lifelong rodent study can be used, integrating the knowledge on the mode of action, pharmacology, and other relevant molecular and functional attributes with the data from other studies (such as the general toxicity). If a positive risk is anticipated, mechanistic studies and mitigation strategies are considered more adequate than the classical 2-year rodent study. In conclusion, although still based on the same principles, the classical animal-based paradigm used for small molecules was adapted for biopharmaceuticals. Also, *in vitro* only strategies may be acceptable to support the introduction of a molecular entity into humans, when the animal species will prove irrelevant [24].

16.4.2 Advanced Therapy Medicinal Products

Following biopharmaceuticals, innovative therapies based on cells (cell therapy medicinal products (CTMPs)), engineered tissues, or gene modulation (gene therapy medicinal products (GTMPs)) are emerging with innovative scientific challenges that are difficult to address based on the current paradigm. In contrast to the small chemical entities, the advanced therapy medicinal products (ATMPs) involve quality, preclinical, and clinical aspects, which cannot be studied in sequence due to high degree of interconnection. For instance, several quality attributes overlap or greatly affect the biological activity of the medicinal product, and therefore need to be discussed and studied also under the preclinical and clinical settings. Examples are the cell migration or proliferation properties in a CTMP, or the integrative properties of a viral vector used for a GTMP. Due to the specific scientific aspect of those products, a new committee, the CAT (Committee for Advanced Therapies), was created at the EMA, to deal with all scientific and regulatory aspects of ATMPs, including the preparation of guidelines and the assessment of marketing authorization applications, together with the CHMP [27]. With regard to the preclinical development plans for ATMPs, based on the recognition that the classical paradigm would not be applicable in many cases, the risk-based approach has been created [28]. This is defined as a strategy aiming to

Table 16.4 Examples of risk factors that can be associated with ATMPs [28].

Cell therapy medicinal products	Gene therapy medicinal products
Origin of cell (autologous or allogeneic)	The potential of the vector for and its extent of chromosomal integration
Ability to migrate from the site of application	The capacity of the vector for latency/reactivation and/or mobilization
Ability to engraft in ectopic sites, proliferate, and differentiate	Its potential for recombination/reassortment and biodistribution to nontarget sites
Ability to initiate immune response	Expression of the therapeutic or any other transgene delivered and duration of expression.
Level of cell manipulation (*in vitro*/*ex vivo* activation/genetic manipulation)	Replication – incompetence or competence of a vector and its capacity to inadvertently replicate after complementation by a respective wild-type or helper virus
Aspects of the manufacturing process	Patient-related risks
Noncellular components	Disease-related risks
Mode of administration	
Ex vivo perfusion, local, systemic	
Duration of exposure (short to permanent)	
Patient-related risks	
Disease-related risks	

determine the extent of quality, nonclinical, and clinical data to include in the Marketing Authorization Application (MAA), in accordance with the respective scientific guidelines, and to justify any deviation from the technical requirements as defined in Annex I, Part IV of Directive 2001/83/EC. The risk-based approach consists of "predictive" rather than "reactive" thinking and can be considered closer to a paradigm shift. Based on the existing/built knowledge of the ATMPs, the safety assessment plan is designed to address and clarify the anticipated risks, taking into consideration different product-related factors (Table 16.4).

The safety assessment of any ATMP is therefore based on risk profiling, defined as a methodological approach to systematically integrate all available information on risks and risk factors, in order to obtain a profile of each individual risk associated with a specific ATMP. Four steps are proposed for risk profiling: (1) identify risks associated with the clinical use of the ATMP; (2) identify product-specific risk factors contributing to each identified risk; (3) map the relevant data for each identified risk factor against each of the identified risks; and (4) conclude on the risk factor–risk relationships. The risk-based approach is flexible, less standardized compared with the classical paradigm, and leads to case-based programs, still under GLP principles. It requires a high knowledge of the new therapeutic entity, obtained in anticipation, and will mostly allow the prediction of adverse effects in a faster and more reliable way, compared with the standard animal tests, for which the basis is mostly the observation of effects,

not their prediction. A product-specific strategy is often built for those products, under discussion with Regulatory Authorities, for example, through EMA and/or FDA Scientific Advice Procedures [29]. Similarly to biologics, Regulatory Authorities do recognize that animal-based studies can be irrelevant or redundant in some cases for ATMPs, and other approaches can be used to address safety assessment, involving, for example, *in vitro* systems, based on human-derived cells/tissue systems, followed by appropriate, safely conducted, studies in humans [28]. The increased knowledge gathered from cell and gene therapies is helping to build new human cell-based systems (stem cell-derived systems) for studying, for example, the mode of action, potential efficacy, or the toxicity of innovative medicines. A contribution for paradigm modernization is possible.

16.4.3 Nanopharmaceuticals

Nanotechnology-based products have added new challenges to the classical preclinical paradigm. *In silico* systems, including modeling and simulation, using the intended attributes of the nanosystems are used to shape their specific targeting attributes (e.g., target access, tissue and intracellular access, and distribution) and biodisposition including elimination characteristics, subsequently confirmed *in vitro* using imaging techniques for tracking particles intracellularly. By acknowledging that the innovative delivery system or the size of the drug (nano)particles may affect the activity at multiple levels, including potency, local of action, (intra)cellular distribution, and so on [30], it is recognized that innovative tests are needed, which are not necessarily in line with the classical paradigm.

The benefit gained from the full therapeutic potential of nanosized formulations still depends on further basic knowledge on cell and molecular biology at intracellular level, to fully understand how the nanoparticles are presented to organs, cells, and organelles, according to the different mechanisms of intracellular trafficking and its consequences. Toxicological aspects of nanomedicines have been highlighted with focus on long-term toxicity. Carbon nanotubes, quantum dots, and other nonbiodegradable and potentially harmful materials need close attention, whether associated with medicines or diagnostics. A dedicated set of standards is needed in the global regulatory environment for the nanosystems being developed. In fact, some regulatory elements have already been produced, and new dedicated guidance documents, the so-called reflection papers, have been prepared and published by the EMA in collaboration and under interaction with Regulatory Authorities from the United States and Japan, addressing aspects to consider for the development of different types of nanopharmaceuticals, including the preclinical safety characterization. By recognizing the need to produce specific guidance for these products, the Regulatory Authorities implicitly recognize the need for adaptations of the current standard preclinical paradigm to address/predict their bioactivity in humans [31].

Table 16.5 The stepwise approach for nonclinical testing of mAbs [32].

First step	Comparative *in vitro* studies to assess differences in binding or functions
Second step	Decide whether additional *in vivo* nonclinical work is warranted
Third step	Study plan and conduct

In vitro studies: comparative studies including those available from quality-related assays, for example,

- Binding to target antigen(s)
- Binding to Fc gamma receptors (FcγRI, FcγRIIA, FcγRIIB, FcγRIIIA, and FcγRIIIB), FcRn, and complement
- Fab-associated functions (e.g., neutralization, receptor activation, or receptor blockade)
- Fc-associated functions (e.g., ADCC and CDC assays, complement activation)

16.4.4 Biosimilar Medicinal Products

A guideline addressing the preclinical and clinical development of biosimilar mAbs has recently been issued by the European Medicines Agency [32], which deviates from the animal *in vivo* studies for proof of biosimilarity. It is recognized that animal studies to compare the toxicity or the pharmacological activity of the biosimilar and the reference product are not sufficiently powered to identify small differences in specific key attributes. A staged approach focusing first on quality characterization, followed by *in vitro* analysis of the mAb bioactivity on its targets, is more appropriate (Table 16.5). The decision on the need for animal studies depends on the level of (un)certainty reached with those studies, or on other factors, for example, impurities, among others.

The low, if any, relevance of animal studies for proof of biosimilarity, particularly with mAbs, again shows that alternative strategies to the standardized animal-based paradigm are possible and needed, and can be planned in a case-based manner, taking benefit of advances in *in vitro* and *in silico* (modeling) systems, in a prospective rather than reactive manner. Also in this case, the stepwise strategy aligns with the risk-based approach. It proposes that the nonclinical studies are performed based on a previously identified rationale, appropriately justified: *in vitro* followed by *in vivo* if needed, performed before the clinical development, in line with the acceptance by Regulatory Authorities in a case-based manner.

16.4.5 Innovative Small Chemical Entities

Despite that the need for classical preclinical paradigm change or adaptation has been driven by the reduced value of at least some of the inherent established approaches, as was the case for species selection for biopharmaceuticals, exemption for long-term studies or carcinogenicity, no need of genotoxicity testing, and so on, also the development of innovative small chemical entities is raising,

in many cases, supportive evidence for the possibility of using alternatives to the classical paradigm, more prospective, less reactive. Indeed, small molecular entities are increasingly being designed to be specifically interacting with an established target, for which a relevant amount of knowledge has been generated, including its role in the disease, the associated cascades, cascade crosstalks, and so on. Even in situations where the target might be less known, companies devote huge efforts to characterize the cellular cascades associated with the mode of action, which leads to the possibility of anticipating multiple aspects of efficacy and safety. As a consequence of the target specificity, pharmacologically driven adverse effects are also increasingly being identified with small molecules, raising the question of whether many of the principles that have been applied to the biopharmaceuticals, deviating from the classical paradigm for preclinical safety testing, or ATMPs (e.g., the risk-based approach) should also be applied to those new developing molecules/candidates in general. Indeed, as an example, such way of thinking is behind the revision of the ICH S1 topic (see Section 16.1.3). While the understanding of the mode of action in the past may have been at a superficial stage when lead candidates were selected to pursue the development, the current advances in molecular biology and genetics/genomics, associated with the huge technological advances, are making possible to collect a very high amount of information with regard to the cellular effects of drugs, their targets, the cascades directly and indirectly associated with the molecular activity, the crosstalk with other cellular/target cascades, and so on, irrespective of whether they are biologics or small molecules. With this knowledge in place, it is expected that the potential for pharmacology-related effects might be anticipated and/or monitored for their occurrence. For small molecules, the structure-related toxicity is a concern, but the currently existing or "under construction" databases will provide a basis for predicting the toxicity unrelated to the MoA. A prospective approach, rather than a reactive one, may also make sense for most if not all of these molecules. These reflections point toward a risk-based approach-based paradigm to be applied also to small molecules.

16.5 Envisaging a Paradigm Change

16.5.1 The Present

It can be considered that the application of the current paradigm on drug development starts with the lead candidate selection, using screening approaches aiming to reach the most effective, bioavailable, most appropriately biodistributed, and potentially safe molecule. The lead candidate selection involves high-throughput screening assays mostly *in sílico* and *in vitro*, to exclude the molecules for which some unacceptable safety attributes in the context of the intended therapeutic use are identified, such as genotoxicity (in cells and

bacteria), cardiotoxicity (in transfected cells or cardiomyocytes), cytotoxicity (in multiple types of cells such as hepatocytes and others), and embryotoxicity (in zebrafish embryos or embryonic stem cells). Also, exploratory clinical trials, for example, with microdoses (microdose studies) for the characterization of, for example, intended target binding, bioavailability, or any efficacy biomarker can help in the selection process.

Following its selection, the development of the lead candidate is a stepwise process that, after the first set of short-term studies supportive of the phase I clinical trials, includes multiple studies aimed at supporting the safety of the patients involved in clinical trials up to market. The core preclinical development program can be seen as a strategy to respond to a set of questions aiming at characterizing the biological activity and the consequences for the body (pharmacodynamics and toxicity), as well as the body actions over the product (pharmacokinetics and metabolism). The study program is expected to respond, mostly in nonhuman models, among others, to the following questions:

- How is the drug–target interaction?
- How does the target modulation affect the intended disease?
- Is the molecule able to affect any body function relevant for life, for example, the central nervous system or the cardiovascular system, with, for example, proarrhythmic effects?
- Which are the additional targets and what is the consequence of their modulation?
- How does the body work on the molecule? (ADME)
- How and how long does the drug distribute or accumulate in different organs and tissues?
- Is the drug inducing toxicity at any organ or tissue?
- Is the drug toxic for the reproductive function?
- Is the drug carcinogenic?
- Are there any other types of toxicity, for example, over the CNS, the immune system, or the skin and eyes due to light activation?

The human risk is subsequently assessed based on the interpretation of the data collected (systems' responses) taking into account the limitations of the models used for the human situation, the conditions at which the toxic effects were seen, and the conditions of human use, making the balance between the expected benefits and risks. Often, additional studies of mechanistic nature are needed for the understanding of the human relevance of the effects judged as important.

16.5.2 The Basis for a Paradigm Change

The use of the current paradigm for nonclinical safety testing of pharmaceuticals for the half-century has produced an enormous knowledge and experience with the models used, mostly animal-based models, and the understanding and

identification of their value and limitations in predicting human safety and efficacy. The enrichment of the standard paradigm with innovative technologies has not helped to improve the success of dug development. On the other hand, the growing availability and experience with the more sophisticated and advanced technologies used in the screening predevelopment phases, combined with outcomes of paradigm adjustments triggered by innovative therapies (biologics, ATMPs, nanopharmaceuticals, and biosimilars), clearly suggest that an extensive reformulation of the current preclinical development plan is possible. This trend is further reinforced by the impressive output of multiple joint research initiatives in the EU or the United States, involving academia, industry, and regulators, mostly based on *in silico* and *in vitro* predictive systems of human origin. While the use of animal models of disease for potential efficacy testing has proven useful although limited in multiple situations, for example, diabetes, hypertension, some genetically driven diseases, and so on, it is less efficient or even inefficient for other situations, such as CNS diseases, stroke, cancer, and so on. Therefore, additional disease-related models reflecting the human conditions are needed. In addition to new animal models obtained by genetic manipulations, multiple approaches are being developed to provide cell-based systems derived from different human conditions/diseases in cocultures or three-dimensional organotypic cultures still under improvement but promising for the understanding of disease-related aspects and identification of mechanisms for overcoming associated dysfunctions. Also, such human cell-based systems are being considered and developed for toxicity assessment (e.g., STEMBANCC (Stem Cells for Biological Assays of Novel Drugs and Predictive Toxicology)) [15]. The efforts being put forward to obtain multiple lineages of human cells derived from non-embryonic stem cells (the induced pluripotent stem cells) with standardized procedures will be a major advance in this direction. Mostly because these initiatives will allow the access to cell systems derived from different types of patients and from healthy volunteers, target characterization, drug–target interactions in healthy and diseased cells/organotypic systems, cascade crosstalk, and so on will be the basis for the understanding of mode of action and pharmacologically driven safety aspects. The use of modeling approaches will make possible in the future to anticipate the conditions for some predicted effects to occur in humans/patients, and might allow the faster progress from preclinical testing to clinical trials, starting with exploratory studies (ICH M3(R2)), avoiding or drastically reducing the animal testing. Aspects for replacement of animal studies are also considered by, for example, the European Regulatory Authorities [33].

Putting together the advances generated by built-for-purpose strategies in the EU, United States, and other regions of the world, involving academia, industry, and regulators, it can be anticipated that a number of tools are becoming available to be combined in different models to predict the following:

1) *Potential efficacy of candidate molecules:* based on their activity at relevant targets, taking into consideration the target relevance and involvement in the disease (previously studied). Relevant for the identification of

disease-associated targets and other relevant disease-modulated phenomena are the induced pluripotent cell-derived cell systems, collected from different types of patients, which may provide "human" or "humanized" tools to study potential efficacy, in addition or alternatively to animal models of disease.

2) *Pharmacokinetic attributes:* based on, for example, molecular structure and *in vitro* systems addressing, for example, absorption, transport, and metabolism; based on primary and secondary target distributions allowing the prediction of target organs; and based on exploratory clinical trials such as the microdosing studies, anticipating some pharmacological, pharmacokinetic, or even safety attributes in humans at very low doses.
3) *Potential safety aspects:* based on the molecular attributes and structure–activity relationships captured from existing and building databases (for toxicities such as genotoxicity, hepatotoxicity, cardiotoxicity, nephrotoxicity, neurotoxicity, etc.); based on the mode of action (class-related effects), the relevant target(s) binding attributes, and target involvement in multiple cellular, tissue, and organ functions, as known or identified in human cell cultures or other systems; and based on the characterization of the effects *in vitro* (using multiple organotypic systems, for example, mimicking the kidney, lung, liver, heart, CNS, etc.). The integration of the different levels of information obtained *in vitro* and *in silico* using computational tools, including modeling and simulation, is the basis for the systems pharmacology and toxicology, which may form the basis for the potential paradigm shift, allowing a faster progress to human studies.
4) *Preliminary human pharmacokinetic, pharmacodynamic, and safety attributes* in vivo *(exploratory clinical trials):* the transposition of the data obtained in alternative *in silico* and *in vitro* human cell-based test systems into the first-in-human study will/would constitute a critical step for which measures need to be implemented to allow the anticipation of any relevant safety aspect, which might derive from the first application of the investigational molecule. For this purpose, the advances in the imaging and biomarker fields will contribute decisively to the success of the preclinical to clinical translation, irrespective of the preclinical paradigm that might have been followed. Highly sensitive and specific predictive safety biomarkers are needed, and many are being investigated, to enable the anticipation of any emerging toxic events before histopathological changes of the affected organs have occurred, as well as the monitoring of the reversibility. When sufficiently sensitive and specific, these biomarkers will allow the earlier initiation of human investigations of innovative drugs in safe conditions, with early tracking of signals on organ toxicity, before any structural/functional damages have occurred. Hopefully, these new tools, when successfully implemented, will reposition the exploratory human studies in the drug development process, and the need and the usefulness of at least some animal-based preclinical studies, which could be reduced, adapted, or ultimately eliminated.

16.5.3 Vision of a Renewed Paradigm

An innovative paradigm, to be accepted, should provide early anticipation of the lack of efficacy of any developing candidate despite promising nonclinical tests, as well the potential for unacceptable adverse drug reactions.

A new paradigm, where multiple steps/components, not necessarily sequential, could be described, somehow in line with the risk-based approach proposed for ATMPs, appears to be emerging from the ongoing scientific and technological advances, as follows:

1) *Comprehensive characterization of the molecular candidate*:
   - Structural features, physical and chemical attributes.
   - The relevant primary and secondary targets, their functionalities and distribution across human cells and systems of different tissues and organs.
   - The biodisposition of the molecule and attributes of relevant human metabolites using *in vitro* systems and modeling and simulation approaches.
2) *Activity profiling*: anticipation of effects, either wanted or adverse, related to the primary and the secondary pharmacology of the molecule and its relevant metabolites through the integrated information from the bullet points above (qualitative and quantitative).
3) *Risk profiling*:
   a) *General risk prediction*: use of *in silico* tools to anticipate structure-related toxicity profile, for example, genotoxicity, hepatotoxicity, nephrotoxicity, neurotoxicity, cardiotoxicity, dermatotoxicity, or eye toxicity (due to phototoxicity).
   b) *Experimental risk testing:* test the anticipated toxicities based on point 2 in *in vitro* human cell-based systems, from healthy or diseased subjects. Those cells will be under organotypic cultures, interconnected through microfluidic circulation (human organs on a chip in projects such as NCATS's "Re-engineering Translational Sciences") [34].
4) *Translation for human safe starting testing conditions*: integrative treatment of the full data to anticipate the pharmacological dose, toxicological dose, and the safe starting dose for an exploratory first-in-human study (microdosing study), together with identification of which functions/systems to monitor mainly through predictive biomarkers.
5) Further clinical study designs of exploratory nature (as of ICH M3(R2) proposals) might allow to pursue with repeated dosing, still under close monitoring of subjects using appropriate tools, including the appropriate setting of omics markers to screen for the emergence of anticipated toxicities or new ones not predicted.
6) *Further investigations*: any findings from first or subsequent clinical studies would trigger further investigations, experimental or clinical, *in vitro*

(animals not being excluded) to provide the knowledge and the conditions for safe continuation of the clinical investigation. A spectrum of general safety and safe conditions of use of the investigational product should be obtained through this strategy, driving research back and forth from bench to bed to bench. Strategies or new approaches for clinical trials are not to be discussed in this section, but will need to be considered together with any shifts of the nonclinical paradigm.

7) *Specific risk prediction*: questions as posed with the classical nonclinical safety paradigm will still need to be responded, as are the cases for carcinogenicity and reproductive toxicity. Both topics are under discussion with respect to the possible introduction of innovative approaches limiting the burden of animal studies in both situations. For reproductive toxicity, while aspects might be able to be addressed using *in silico* and *in vitro* systems, for example, germ cells, embryos, and embryonic stem cells, together with hormonal effects of drugs, among others, it is recognized that teratogenicity is a very sensitive aspect that will take time to be resolved through alternative approaches to those currently in place, based on two animal species, the rat and the rabbit. However, for cases where such species prove irrelevant, for example, the case of monoclonal antibodies for which the nonhuman primate is often the only relevant species, the study design has been changed and adapted, as stated in ICH S6(R1) guideline. In conclusion, based on the progress currently in place, it can be expected that in the future the preclinical safety paradigm will be undergoing extensive changes, moving from a "reactive", sequential, animal-based approach to a "proactive", *in vitro* and *in sílico* human-based approach, in a more dynamic "bench to bed to bench . . ." strategy, hopefully leading to a significant reduction in the (late) attrition rates, and to safe and faster access of patients to more efficacious medicines.

16.6 Regulatory Actions Needed to Shift the Animal-Based Paradigm

In this chapter, it has been attempted to provide a historical overview of the preclinical safety testing paradigm, from its implementation until its current format and content, emphasizing the scientific advances that have been occurring and their potential impact on future modifications toward a possible shift of the current design.

However, the paradigm is based on a large number of concepts and studies heavily regulated by national or regional laws (Europe, United States, Japan, ICH, etc.) through a large number of guidelines and regulations that are mandatory. The ultimate goal of the process is to protect the health of subjects in clinical trials and of treated patients. In this context, any paradigm change will need to be supported by data showing that, compared with the existing one, no additional risks are introduced with the new processes.

Most regulatory changes and updates in the preclinical setting have been achieved based on studies where innovative strategies are compared with the existing ones, often implying the validation of new tests against the existing ones, often animal-based tests. This process "perpetuates" the positioning of the animal testing, if the approach is not changed. Although not extensive, in some areas *in vitro* tests have been recognized as prone to replace *in vivo* studies, and were implemented in the corresponding guidelines. An example is the use of hERG channel assay as a predictor of proarrhythmogenic potential. Recognizing that multiple *in vitro* test systems are under development and others are being used by pharmaceutical companies in their screening predevelopment programs, a guideline on 3Rs has been produced by the EMA, which is currently being updated [35]. However, if a paradigm shift away from the animal-based strategies is developed, the full strategy will have to be tested against the existing one, in global, and not in a "study-by-study" manner, due to the large differences in the approaches. For instance, the prediction of general toxicity based on the new paradigm will incorporate a multitude of approaches (e.g., pharmacodynamically based approaches and *in vitro* and *in silico* human-based approaches), which will not be possible to compare with each single outcome of the full repeated-dose toxicity study. Indeed, the latter will mostly provide effects in animals, while the former will be mostly based on human systems. Therefore, any new paradigm validation will need to be based on full comparative data derived from "old" and "new" paradigms. One way to achieve this goal could be by using available data from already fully developed molecules for which experience is available, and introducing them in innovative paradigm, using a blinded approach, where the researchers would not know the testing molecule. A stepwise approach starting with the testing program to support a first-in-human study could be a starting point. The Regulatory Authorities (together with the pharmaceutical industry) are experienced with processes such as certification of predictive safety biomarkers, as models for this exercise. New or adapted guidance will be needed to allow the harmonization and standardization of the different set of (new) studies. A new paradigm will have to be as trusted as the existing one by all stakeholders, regulators, industry, patients, and the society in general. The ICH will most possibly be the ideal forum for undertaking and processing these actions, in line with other ongoing initiatives such as the S1 topic update. In conclusion, with a note of optimism, it is worth mentioning that, although no important changes have been introduced in the general format of the preclinical paradigm for drug development, some updates in the study requirements have been introduced in line with modernizations triggered by innovative techniques, some of which use alternative approaches to the classical animal studies. These updates are substantiated in guidance changes, showing the awareness and willingness of Regulatory Authorities to adapt the current paradigm, provided that the new strategies bring advantages relative to the previously existing ones. By changing the current paradigm into an innovative, more efficient one, it is expected that the drug development process in general will become more fast and efficient and the cost involved will be

considerably reduced due to a substantial reduction in the late attrition rates leading to a global benefit for public health.

References

1 FDA (1938) Significant Dates in US Food and Drug Law History, Available at http://www.fda.gov (last accessed February 2014).
2 Rice, E. (2007) Dr. Frances Kelsey: Turning the Thalidomide Tragedy into Food and Drug Administration Reform. Research paper.
3 Greek, R., Shanks, N., and Rice, M.J. (2011) The history and implications of testing thalidomide on animals. *J. Philos. Sci. Law*, **11**, 1–32.
4 Schardein, J.L. (1976) *Drugs as Teratogens*, CRC Press.
5 Lehman, A.J. *et al.* (1949) Procedures for the appraisal of the toxicity of chemicals in foods. *Food Drug Cosmet. Law Q.*, **4** (3), 412–434.
6 Jacobs, A.C. and Hatfield, K.P. (2012) History of chronic toxicity and animal carcinogenicity studies for pharmaceuticals. *Vet. Pathol.*, **50** (2), 324–333.
7 Rägo, L. and Santoso, B. (2008) Drug regulation: history, present and future, in *Drug Benefits and Risks: International Textbook of Clinical Pharmacology*, revised 2nd edn (eds C.J. van Boxtel, B. Santoso, and I.R. Edwards), IOS Press and Uppsala Monitoring Centre.
8 European Commission (1965) Council Directive 65/65/EEC of 26 January 1965 on the approximation of provisions laid down by law, regulation or administrative action relating to medicinal products. *Off. J. L*, **22**, 369.
9 European Commission (1975) Council Directive 75/318/EEC of 20 May 1975 on the approximation of the laws of Member States relating to analytical, pharmaco-toxicological and clinical standards and protocols in respect of the testing of proprietary medicinal products. *Off. J. L*, **147**, 1–12.
10 European Commission (1987) Council Directive 87/22/EEC of 22 December 1986 on the approximation of national measures relating to the placing on the market of high-technology medicinal products, particularly those derived from biotechnology. *Off. J. L*, **015**, 0038–0041.
11 European Commission (1975) Second Council Directive 75/319/EEC of 20 May 1975 on the approximation of provisions laid down by law, regulation or administrative action relating to proprietary medicinal products. *Off. J. L*, **147**, 0013–0022.
12 European Commission (1995) Commission Regulation (EC) No 540/95 of 10 March 1995 laying down the arrangements for reporting suspected unexpected adverse reactions which are not serious, whether arising in the Community or in a third country, to medicinal products for human or veterinary use authorized in accordance with the provisions of Council Regulation (EEC) No 2309/93. *Off. J. L*, **055**, 0005–0006.
13 European Commission (1989) The Rules Governing Medicinal Products in the European Community. Catalogue No. CB-55-89-706-EN-C, Office for Official Publications of the European Communities, Luxembourg.
14 ICH, About ICH: History. Available at www.ich.org (last accessed February 2014).
15 Innovative Medicines Initiative, Ongoing Projects: MARCAR. Available at www.imi.europa.eu (last accessed February 2014).
16 EMA (1995) ICH Guideline M3(R2) on Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorisation for Pharmaceuticals. EMA/CPMP/ICH/286/1995. Available at www.ema.europa.eu (last accessed February 2014).
17 Kola, I. and Landis, J. (2004) A survey of pharmaceutical companies comparing reasons for attrition. *Nat. Rev. Drug Discov.*, **3**, 711–715.

18 Arrowsmith, J. and Miller, P. (2013) Phase I and phase II attrition rates 2011–2012. *Nat. Rev. Drug Discov.*, **12**, 569.

19 Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W., Koen Van Deun, M., Smith, P., Berger, B., and Heller, A. (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul. Toxicol. Pharmacol.*, **32**, 56–67.

20 Innovative Medicines Initiative (2008) The Innovative Medicines Initiative (IMI) Research Agenda. Creating Biomedical R&D Leadership for Europe to Benefit Patients and Society, Version 2.0.

21 Goodsaid, F.M., Frueh, F.W., and Mattes, W. (2007) The Predictive Safety Testing Consortium: a synthesis of the goals, challenges and accomplishments of the critical path. *Drug Discov. Today Technol.*, **4** (2), 47–50.

22 Sistare, F.D., Dieterle, F., Troth, S., Holder, D.J., Gerhold, D., Andrews-Cleavenger, D., Baer, W., Betton, G., Bounous, D., Carl, K., Collins, N., Goering, P., Goodsaid, F., Gu, Y.-Z., Guilpin, V., Harpur, E., Hassan, A., Jacobson-Kram, D., Kasper, P., Laurie, D., Silva-Lima, B., Maciulaitis, R., Mattes, W., Maurer, G., Obert, L.A., Ozer, J., Papaluca-Amati, M., Phillips, J.A., Pinches, M., Schipper, M.J., Thompson, K.L., Vamvakas, S., Vidal, J.M., Vonderscher, J., Walker, E., Webb, C., and Yu, Y. (2010) Towards consensus practices to qualify safety biomarkers for use in early drug development. *Nat. Biotechnol.*, **25**, 446–454.

23 Dieterle, F., Sistare, F., Goodsaid, F., Papaluca, M., Ozer, J.S., Webb, C.P., Baer, W., Senagore, A., Schipper, M.J., Vonderscher, J., Sultana, S., Gerhold, D.L., Phillips, J.A., Maurer, G., Carl, K., Laurie, D., Harpur, E., Sonee, M., Ennulat, D., Holder, D., Andrews-Cleavenger, D., Gu, Y.-Z., Thompson, K.L., Goering, P.L., Vidal, J.M., Abadie, E., Maciulaitis, R., Jacobson-Kram, D., Defelice, A.F., Hausner, E.A., Blank, M., Thompson, A., Harlow, P., Throckmorton, D., Xiao, S., Xu, N., Taylor, W., Vamvakas, S., Flamion, B., Lima, B.S., Kasper, P., Pasanen, M., Prasad, K., Troth, S., Bounous, D., Robinson-Gravatt, D., Betton, G., Davis, M.A., Akunda, J., McDuffie, J.E., Suter, L., Obert, L., Guffroy, M., Pinches, M., Jayadev, S., Blomme, E.A., Beushausen, S.A., Barlow, V.G., Collins, N., Waring, J., Honor, D., Snook, S., Lee, J., Rossi, P., Walker, E., and Mattes, W. (2010) Renal biomarker qualification submission: a dialog between the FDA-EMEA and Predictive Safety Testing Consortium. *Nat. Biotechnol.*, **28**, 455–462.

24 EMA (1998) ICH Guideline S6(R1): Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals. EMA/CHMP/ICH/731268/1998. Available at www.ema.europa.eu (last accessed February 2014).

25 van Meer, P.J., Kooijman, M., van der Laan, J.W., Moors, E.H., and Schellekens, H. (2013) The value of non-human primates in the development of monoclonal antibodies. *Nat. Biotechnol.*, **31**, 882–883.

26 EMA (2007) Guideline on Strategies to Identify and Mitigate Risks for First-in-Human Clinical Trials with Investigational Medicinal Products. EMEA/CHMP/SWP/28367/07. Available at www.ema.europa.eu (last accessed February 2014).

27 Schneider, C.K., Salmikangas, P., Jilma, B., Flamion, B., Todorova, L.R., Paphitou, A., Haunerova, I., Maimets, T., Trouvin, J.H., Flory, E., Tsiftsoglou, A., Sarkadi, B., Gudmundsson, K., O'Donovan, M., Migliaccio, G., Ancāns, J., Maciulaitis, R., Robert, J.L., Samuel, A., Ovelgönne, J.H., Hystad, M., Fal, A.M., Lima, B.S., Moraru, A.S., Turcáni, P., Zorec, R., Ruiz, S., Akerblom, L., Narayanan, G., Kent, A., Bignami, F., Dickson, J.G., Niederwieser, D., Figuerola-Santos, M.A., Reischl, I.G., Beuneu, C., Georgiev, R., Vassiliou, M., Pychova, A., Clausen, M., Methuen, T., Lucas, S., Schüssler-Lenz, M., Kokkas, V., Buzás, Z., MacAleenan, N., Galli, M.C., Linē, A., Gulbinovic, J., Berchem, G., Fraczek, M., Menezes-Ferreira, M., Vilceanu, N., Hrubisko, M., Marinko, P., Timón, M., Cheng, W., Crosbie, G.A., Meade, N., di Paola, M.L., VandenDriessche, T., Ljungman, P., D'Apote, L., Oliver-Diaz, O., Büttel, I., and Celis, P. (Committee for Advanced

Therapies (CAT); CAT Scientific Secretariat) (2010) Challenges with advanced therapy medicinal products and how to meet them. *Nat. Rev. Drug Discov.*, **3**, 195–201.

28 EMA (2011) Guideline on the Risk-Based Approach According to Annex I, Part IV of Directive 2001/83/EC Applied to Advanced Therapy Medicinal Products. EMA/CAT/CPWP/686637/2011. Available at www.ema.europa.eu (last accessed February 2014).

29 EMA (2010) European Medicines Agency Guidance for Companies Requesting Scientific Advice and Protocol Assistance. EMEA-H-4260-01 Rev. 6. Available at www.ema.europa.eu (last accessed February 2014).

30 McNeil (2009) Nanoparticle therapeutics: a personal perspective. *Nanomed. Nanobiotechnol.*, **1**, 264–271.

31 Ehmann, F., Sakai-Kato, K., Duncan, R., Hernán Pérez de la Ossa, D., Pita, R., Vidal, J.M., Kohli, A., Tothfalusi, L., Sanh, A., Tinton, S., Robert, J.L., Silva Lima, B., and Amati, M.P. (2013) Next-generation nanomedicines and nanosimilars: EU regulators' initiatives relating to the development and evaluation of nanomedicines. *Nanomedicine*, **8** (5), 849–856.

32 EMA (2005) Guideline on Similar Biological Medicinal Products Containing Monoclonal Antibodies: Non-Clinical and Clinical Issues. EMA/CHMP/BMWP/403543/2010. Available at www.ema.europa.eu (last accessed February 2014).

33 Manolis, E., Rohou, S., Hemmings, R., Salmonson, T., Karlsson, M., and Milligan, P.A. (2013) The role of modeling and simulation in development and registration of medicinal products: output from the EFPIA/EMA Modeling and Simulation Workshop. *CPT Pharmacometrics Syst. Pharmacol.*, **2**, e31.

34 Colins, F.S. (2011) Reengineering translational science: the time is right. *Sci. Transl. Med.*, **3** (90), 90cm17.

35 EMA (2011) Concept Paper on the Need for Revision of the Position on the Replacement of Animal Studies by *In Vitro* Models. EMA/CHMP/SWP/169839/2011. Available at www.ema.europa.eu (last accessed February 2014).

Index

Predictive Toxicology: From Vision to Reality, First Edition. Edited by Friedlieb Pfannkuch and Laura Suter-Dick.

h