Springer Protocols

Usha Kasid
Robert Clarke   *Editors*

Cancer Gene
Networks

Humana Press

METHODS IN MOLECULAR BIOLOGY

For further volumes:
http://www.springer.com/series/7651

Cancer Gene Networks

Edited by

Usha Kasid and Robert Clarke

Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, District of Columbia, USA

Humana Press

Editors
Usha Kasid
Lombardi Comprehensive Cancer Center
Georgetown University Medical Center
Washington
District of Columbia
USA

Robert Clarke
Lombardi Comprehensive Cancer Center
Georgetown University Medical Center
Washington
District of Columbia
USA

Preface

In recent years, the application of systems biology to increase our understanding of cancer has led to the integration of a variety of the state-of-the-art high-throughput and multidisciplinary tools and workflows. These advances have begun to reveal further some of the complexity, and the dynamic nature, of gene networks and signaling pathways that vary from tumor to tumor and also during the course of the disease. In this volume, a wide range of topics contributed by leading experts in the fields of cancer systems biology and therapeutics, medical bioinformatics, and computational biology offer both technical and methodological guidance and expert insights into many aspects associated with the study of cancer gene networks. Hence, this book provides a valuable and timely resource for a broad audience with interests in basic and translational cancer biology, cancer drug development, as well as in the practice of oncology.

Washington, DC, USA   *Usha Kasid*
                                                                               Robert Clarke

Contents

Contributors

MARTINA BAZZARO • *Masonic Cancer Center and Department of Obstetrics, Gynecology and Women's Health, University of Minnesota Twin Cities, Minneapolis, MN, USA*

AARON BUECHLEIN • *Indiana University Center for Genomics and Bioinformatics, Bloomington, IN, USA*

DEBYANI CHAKRAVARTY • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

JEREMY CHIEN • *University of Kansas Medical Center, Kansas City, KS, USA*

ROBERT CLARKE • *Department of Oncology, Georgetown Lombardi Comprehensive Cancer Center, Washington, DC, USA*

TIMOTHY F. DAY • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

STEFAN DÜBEL • *Technische Universität Braunschweig, Institute of Biochemistry, Biotechnology and Bioinformatics, Braunschweig, Germany*

OFER EIDELMAN • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

FANG FANG • *Medical Sciences Program, Indiana University School of Medicine, Bloomington, IN, USA*

HANNAH C. FELDMAN • *Department of Chemistry, University of Washington, Seattle, WA, USA; Department of Biochemistry, University of Washington, Seattle, WA, USA*

JAMES B. FORD • *Indiana University Center for Genomics and Bioinformatics, Bloomington, IN, USA*

DAVID FRANKHOUSER • *Department of Internal Medicine, OSUCCC-Illumina Core, Columbus, OH, USA*

ROBERT A. GATENBY • *Department of Diagnostic Radiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA; Department of Integrative Mathematical Oncology, H. Lee Moffitt Cancer Center, Tampa, FL, USA*

KIERAN F. GEOGHEGAN • *Structural Biology and Biophysics, Pfizer Worldwide Research, Groton, CT, USA*

MAJDA HAZNADAR • *Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA*

TIM H. HUANG • *Department of Molecular Medicine, University of Texas Health Science Center, San Antonio, TX, USA*

DANWEI HUANGFU • *Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA*

VICTOR X. JIN • *Department of Molecular Medicine, University of Texas Health Science Center, San Antonio, TX, USA*

CATHERINE JOZWIK • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

USHA N. KASID • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

JOO YEUN KIM • *Department of Diagnostic Radiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA; Department of Integrative Mathematical Oncology, H. Lee Moffitt Cancer Center, Tampa, FL, USA*

KARL KROLL • *Department of Internal Medicine, OSUCCC-Illumina Core, Columbus, OH, USA*

PAOLA LECCA • *Department of Mathematics, University of Trento, Trento, Italy; Senior Member of Association for Computing Machinery, New York, NY, USA*

WOO Y. LEE • *Department of Chemical Engineering and Materials Science, Stevens Institute of Technology, Hoboken, NJ, USA*

XIMENA LEIGHTON • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

GUORUI LI • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA*

XIN LI • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

YUNLONG LIU • *Center for Computation Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA*

DUSTIN J. MALY • *Department of Chemistry, University of Washington, Seattle, WA, USA; Department of Biochemistry, University of Washington, Seattle, WA, USA*

ANDREA L.J. MARSCHALL • *Technische Universität Braunschweig, Institute of Biochemistry, Biotechnology and Bioinformatics, Braunschweig, Germany; Department of Systems Immunology and Braunschweig Integrated Centre of Systems Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany*

DANIELA MATEI • *Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA*

EWY A. MATHÉ • *Biomedical Informatics Department, College of Medicine, Ohio State University, Columbus, OH, USA*

RAJSHREE R. MEWANI • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

SCOTT MICHAELS • *Indiana University Center for Genomics and Bioinformatics, Bloomington, IN, USA*

DAVID F.B. MILLER • *Medical Sciences Program, Indiana University School of Medicine, Bloomington, IN, USA*

ASHLEY MOONEYHAM • *Masonic Cancer Center and Department of Obstetrics, Gynecology and Women's Health, University of Minnesota Twin Cities, Minneapolis, MN, USA*

KENNETH P. NEPHEW • *Medical Sciences Program, Indiana University School of Medicine, Bloomington, IN, USA; Department of Cellular and Integrative Physiology, Indiana University School of Medicine, Indianapolis, IN, USA; Department of Obstetrics and Gynecology, Indiana University School of Medicine, Indianapolis, IN, USA*

CHUONG NGUYEN • *Structural Biology and Biophysics, Pfizer Worldwide Research, Groton, CT, USA*

HARVEY B. POLLARD • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

ANGELA RE • *Laboratory of Computational Oncology, Centre for Integrative Biology, University of Trento, Trento, Italy*

HABTOM RESSOM • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

WILLIAM RITCHIE • *CNRS, UMR 5203, Montpellier, France*

DOUGLAS B. RUSCH • *Indiana University Center for Genomics and Bioinformatics, Bloomington, IN, USA*

HUIDONG SHI • *Department of Biochemistry and Molecular Biology, Georgia Regents University, Augusta, GA, USA*

MEERA SRIVASTAVA • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

JOSHUA STARR • *Department of Anatomy, Physiology and Genetics, Institute for Molecular Medicine, Center for Medical Proteomics, Uniformed Services University School of Medicine, Bethesda, MD, USA*

CAMERON STUMP • *Department of Internal Medicine, OSUCCC-Illumina Core, Columbus, OH, USA*

PAIGE STUMP • *Department of Internal Medicine, OSUCCC-Illumina Core, Columbus, OH, USA*

HAIXU TANG • *Indiana University Center for Genomics and Bioinformatics, Bloomington, IN, USA*

NIPUN VERMA • *Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA; Weill Graduate School of Medical Sciences at Cornell University/The Rockefeller University/Sloan Kettering Institute Tri-Institutional M.D.-Ph.D. Program, New York, NY, USA*

JIANBO WANG • *Department of Molecular Medicine, University of Texas Health Science Center, San Antonio, TX, USA*

GRAHAM M. WEST • *Structural Biology and Biophysics, Pfizer Worldwide Research, Groton, CT, USA*

PEARLLY YAN • *Department of Internal Medicine, OSUCCC-Illumina Core, Columbus, OH, USA*

ZHENQING YE • *Department of Molecular Medicine, University of Texas Health Science Center, San Antonio, TX, USA*

LIBO YUAN • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA*

CONGCONG ZHANG • *Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, Frankfurt, Germany*

WENTING ZHANG • *Department of Chemical Engineering and Materials Science, Stevens Institute of Technology, Hoboken, NJ, USA*

ZENGRONG ZHU • *Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA*

ZHIHAO ZHUANG • *Department of Chemistry and Biochemistry, University of Delaware, Newark, DE, USA*

JENNY ZILBERBERG • *John Theurer Cancer Center, Hackensack University Medical Center, Hackensack, NJ, USA*

XIAOJUN ZOU • *Georgetown-Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA*

Chapter 1

Introduction: Cancer Gene Networks

Robert Clarke

Abstract

Constructing, evaluating, and interpreting gene networks generally sits within the broader field of systems biology, which continues to emerge rapidly, particular with respect to its application to understanding the complexity of signaling in the context of cancer biology. For the purposes of this volume, we take a broad definition of systems biology. Considering an organism or disease within an organism as a system, systems biology is the study of the integrated and coordinated interactions of the network(s) of genes, their variants both natural and mutated (e.g., polymorphisms, rearrangements, alternate splicing, mutations), their proteins and isoforms, and the organic and inorganic molecules with which they interact, to execute the biochemical reactions (e.g., as enzymes, substrates, products) that reflect the function of that system. Central to systems biology, and perhaps the only approach that can effectively manage the complexity of such systems, is the building of quantitative multiscale predictive models. The predictions of the models can vary substantially depending on the nature of the model and its inputoutput relationships. For example, a model may predict the outcome of a specific molecular reaction(s), a cellular phenotype (e.g., alive, dead, growth arrest, proliferation, and motility), a change in the respective prevalence of cell or subpopulations, a patient or patient subgroup outcome(s). Such models necessarily require computers. Computational modeling can be thought of as using machine learning and related tools to integrate the very high dimensional data generated from modern, high throughput omics technologies including genomics (next generation sequencing), transcriptomics (gene expression microarrays; RNAseq), metabolomics and proteomics (ultra high performance liquid chromatography, mass spectrometry), and "subomic" technologies to study the kinome, methylome, and others. Mathematical modeling can be thought of as the use of ordinary differential equations and related tools to create dynamic, semi-mechanistic models of low dimensional data including gene/protein signaling as a function of time/dose. More recently, the integration of imaging technologies into predictive multiscale modeling has begun to extend further the scales across which data can be obtained and used to gain insight into system function.

There are several goals for predictive multiscale modeling including the more academic pursuit of understanding how the system or local feature thereof is regulated or functions, to the more practical or translational goals of identifying predictive (selecting which patient should receive which drug/therapy) or prognostic (disease progress and outcome in an individual patient) biomarkers and/or identifying network vulnerabilities that represent potential targets for therapeutic benefit with existing drugs (including drug repurposing) or for the development of new drugs. These various goals are not necessarily mutually exclusive or inclusive.

Within this volume, readers will find examples of many of the activities noted above. Each chapter contains practical and/or methodological insights to guide readers in the design and interpretation of their own and published work.

Key words Systems biology, Cancer gene networks, Quantitative multiscale predictive models, Computational modeling, Mathematical modeling, High throughput omics, Subomics

Constructing, evaluating, and interpreting gene networks generally sits within the broader field of systems biology, which continues to emerge rapidly, particular with respect to its application to understanding the complexity of signaling in the context of cancer biology. For the purposes of this volume, we take a broad definition of systems biology. Considering an organism or disease within an organism as a system, systems biology is the study of the integrated and coordinated interactions of the network(s) of genes, their variants both natural and mutated (e.g., polymorphisms, rearrangements, alternate splicing, mutations), their proteins and isoforms, and the organic and inorganic molecules with which they interact, to execute the biochemical reactions (e.g., as enzymes, substrates, products) that reflect the function of that system. Central to systems biology, and perhaps the only approach that can effectively manage the complexity of such systems, is the building of quantitative multiscale predictive models. The predictions of the models can vary substantially depending on the nature of the model and its input-output relationships. For example, a model may predict the outcome of a specific molecular reaction(s), a cellular phenotype (e.g., alive, dead, growth arrest, proliferation, and motility), a change in the respective prevalence of cell or subpopulations, a patient or patient subgroup outcome(s). Such models necessarily require computers. Computational modeling can be thought of as using machine learning and related tools to integrate the very high dimensional data generated from modern, high throughput omics technologies including genomics (next generation sequencing), transcriptomics (gene expression microarrays; RNAseq), metabolomics and proteomics (ultra high performance liquid chromatography, mass spectrometry), and "subomic" technologies to study the kinome, methylome, and others. Mathematical modeling can be thought of as the use of ordinary differential equations and related tools to create dynamic, semi-mechanistic models of low dimensional data including gene/protein signaling as a function of time/dose. More recently, the integration of imaging technologies into predictive multiscale modeling has begun to extend further the scales across which data can be obtained and used to gain insight into system function.

There are several goals for predictive multiscale modeling including the more academic pursuit of understanding how the system or local feature thereof is regulated or functions, to the more practical or translational goals of identifying predictive (selecting which patient should receive which drug/therapy) or prognostic (disease progress and outcome in an individual patient) biomarkers and/or identifying network vulnerabilities that represent potential targets for therapeutic benefit with existing drugs (including drug repurposing) or for the development of new drugs. These various goals are not necessarily mutually exclusive or inclusive.

Within this volume, readers will find examples of many of the activities noted above. Each chapter contains practical and/or methodological insights to guide readers in the design and interpretation of their own and published work. As a guide, given below is a brief overview of each chapter for readers.

Lecca and Re [1] provide an excellent introduction to the development and use of network-oriented approaches in the context of systems pharmacology for target identification and drug combination optimization. The context dependence of network topology and intrinsic robustness are discussed as are different approaches to explore this topology and identify network vulnerabilities. The determination of highly connected nodes (often called hubs) is considered also in view of node centrality metrics such as betweenness centrality, bridging centrality, and vibrational centrality. The ability to find and optimize effective anticancer drug combinations requires a greater focus on the combination than on the individual drugs. Two general approaches are presented, a non-model based inference of drug–target networks and a network vulnerability analysis applying vibrational centrality. The authors compare node strength estimated from drug–gene network and node centrality in a protein–protein interaction (PPI) network. An understanding of network topology as it relates to cell function can enable optimization of existing therapies and the identification of new combinations or targets.

Tumor heterogeneity is a major problem in the development of drug resistance and further challenges the building of adequate network models that capture all of the vulnerabilities in cancer cells. Kim and Gatenby [2] explore the development of cellular and molecular heterogeneity in the context of Darwinian dynamics. Each tumor represents an ecosystem comprising multiple local habitats of tumors cells living within unique niches. These niches interact with each other and with locally acting environmental selection pressures that drive the dynamical properties of the niches and ecosystem, representing the Habitat Concept of intratumoral heterogeneity. These authors propose that imaging technologies can identify some of these niches and their features, for example, regional variations as affected by immediate anatomical constraints, in much the same way that satellite images can identify geographic localities and their features. Thus, heterogeneity is not driven solely by genetic mutations but also by potentially adaptive responses to factors operating within the microenvironment. The use of such technologies as magnetic resonance imaging (MRI) or images from computed tomography (CT) scans can provide additional information for studying spatial variations reflecting the dynamic properties of the heterogeneity that exists within most solid tumors. The authors hypothesize that MRI and CT scans can identify structures, which can also be monitored over time, and reveal

regional variations within a tumor that reflect the underlying phenotypic and genotypic subpopulations. Such data can create a "species map" of the ecology of a tumor. This approach is termed "radiomics," where cross-sectional imaging provides the data for the development of quantitative algorithms for the mining of conventional clinical images. The chapter provides insight into methods in landscape ecology, where intratumoral heterogeneity in MRI images can be interpreted as representing habitats within each tumor (ecosystem).

Human pluripotent stem cells (hPSCs) offer a powerful tool to study the role of individual genes in development and progression to malignancy. These cells uniquely exhibit the ability to both differentiate into any adult human cell type and to replicate without limit. While the ability to genetically modify hPSCs is mechanistically powerful, they have a low efficiency for transfection and a low rate of spontaneous homologous recombination. Moreover, the need to use a drug-resistance cassette for clonal selection, which can affect the expression of adjacent genes including the reporter gene, usually requires removal of the cassette. Verma et al. [3] provide a detailed protocol for the use of the CRISPR/Cas technology to simply the process for generating fluorescent reporter and epitope tagged hPSCs. The protocol describes several methods including those for transfecting cells, validating established clones, designing donor ssDNA, sequence verifying targeted clones, and using immunoprecipitation and western hybridization to validate epitope tagged lines. The knock-in strategy described could also be used for gene knock-outs. Thus, the procedures described may have utility in a variety of experimental designs including gene overexpression studies and for cell lineage tracing.

RNAseq has become one of the most popular tools for exploring concurrently both expression levels and RNA sequence data. However, chemical modification of bases due to fixation, often from the use of formalin, or degradation can affect sequence and alter apparent abundance, particularly for lower abundance RNA species. These effects are challenging for RNAseq studies that can depend on adequate length reads and abundance for accurate sequence determination and expression level estimates. To address these concerns, Miller et al. [4] describe the use of restructured adapter sequences for library construction with a method applies duplex-specific nuclease to eliminate ribosomal RNAs and retain the RNA species of interest even when the initial input RNA sample is of relatively low quantity. The method, which is presented in detail, can be applied to multiple RNAseq platforms including Illumina HiSeq, MiSeq, and NextSeq.

In mammals, exon-skipping is the most prevalent mechanism for generating alternate spliced RNAs. Several quite effective methods have been described to identify exon skipping but these can

miss some critical events. Wang et al. [5] address the issue of optimally comparing exon-skipping, alternative splicing, transcription and epigenesist through a novel computational work flow that implements a graph-based exon-skipping scanner approach (GESS). Their tool can capture de novo exon-skipping using raw RNASeq data and does not require knowledge of gene annotation. GESS builds a splice-site-link graph, walks across the graph iteratively to map those subgraphs that exhibit a pattern likely reflecting an exon-skipping event. The Mixture of Isoforms (MISO) model is applied to help determine the dominant transcript(s). The authors show the utility of the method in two studies, one of cells of lymphoid origin and one in data from prostate cancer cells.

The tumor microenvironment is widely implicated as a critical contributor to the development and progression of neoplasia and to the ability of tumors to resist the antineoplastic activity of multiple therapeutic interventions. Zhang et al. [6] provide a detailed description of their approaches to modeling the microenvironment of multiple myeloma using two prototype perfusion culture devices. To capture the three-dimensional features of the solid tumor microenvironment, the first device uses an osteoblast-derived tissue scaffold within a microfluidic stricture that can be perfused as appropriate for each study design. The use of a glass slide format makes the device suitable for several experimental designs including performing time-lapse microscopy. For higher throughput, the second approach takes a 96-well format in a modified perfusion culture device where microenvironments can be reconstructed using human and murine cells. This device allows for the preserving of primary cells while providing a high-throughput format for experimentation that is amenable to use on standard plate readers that use standard 96-well tissue culture plates.

Noncoding RNAs are now understood to play critical roles in gene regulation. Micro-RNAs were among the first of these to be identified and there remains considerable interest in more accurately predicting their regulated targets in cancer cells. Most predictive algorithms tend to overfit and predict multiple targets likely reflecting an inflated Type 1 error (false positive). To address this problem Ritchie [7] presents a method that refines predictions and constrains the Type 1 error to generate more robust predictions of miRNA target genes, likely at the cost of an acceptable inflation of the Type 2 error. Three strategies are taken to achieve this overall goal: (1) multi-targeting leverages the observation that the same gene is often targeted multiple times by the same miRNA; (2) widely available gene expression array data supports identification of differentially regulated target genes; (3) integration of information from databases that contain experimental validation of specific targets further supports target identification.

Proteins are the most common molecules targeted by drug and applying novel proteomic technologies to identify new targets has increased in recent years. Nguyen et al. [8] take readers through some of the more recently developed and/or more widely used chemoproteomic methods for drug discovery. Rather than recapitulate much of what has been reviewed by others, the authors focus on some of the more successful methods and some of the more recent approaches that appear to offer great potential to support future discovery. Stability proteomic approaches address experimental designs where the goal is to identify proteins that become thermodynamically unstable or stable once they bind the drug of choice. These methods include cellular thermal shift assay (CETSA), stability of proteins from rates of oxidation (SPROX), and drug affinity-responsive target stability (DARTS). The authors also discuss the role of kinobeads to explore the relative affinities of different enzymes and inhibitors.

The profiling of thousands of endogenous metabolites and small molecules in cells, tissues and fluids including blood and urine can provide insight into function and the nature and consequences of exposures to factors in the environment and lifestyle. The relative ease with which blood and urine can be collected can allow for the analysis of changes over time associated with disease, treatment, response, or other interventions. For cells in culture, changes in metabolism that occur in seconds, minutes, or hours can also be followed and metabolic flux through established pathways can be studied readily. Haznadar and Mathé [9] address several issues in the rapidly developing field of metabolomics with a primary focus on analytical methods and study designs. The authors compare the two primary technological approaches: nuclear magnetic resonance (NMR) and the mass spectrometry (MS)-based approaches that are usually linked to analysis of materials following separation by either liquid or gas chromatography. These are complementary tools since NMR is appropriate mostly for known, high abundance molecules, whereas the MS-based approaches have greater sensitivity and can be more effectively used for profiling and analysis of lower abundance molecules. The authors discuss lipidomics, sample processing, quantification, use of databases to identify molecules from their mass–charge properties, study design, and controls.

Delivery of proteins into cells has remained difficult, especially for larger proteins like antibodies. While there have been many reportedly successful attempts to do so, few methods have included unambiguous validation of delivery into the cytosol. This issues is clearly defined by Marschall et al. [10], who also note the expression of antibodies from transfected cDNAs frequently produce proteins that are incompletely or incorrectly folded if expressed in the cytosol, and so do not function appropriately. Their text includes careful consideration of the most appropriate controls and they present two independent strategies for critical evaluation

of antibody delivery efficiency. A more widespread application of the optimal method allows investigators to deliver antibodies to cells for modulating cell signalling, e.g. by targeting individual phosphorylation sites of cellular receptors.

Genomics and transcriptomics have garnered most of the recent attention. However, it is more often the protein products of these nucleic acids, and their posttranslational modifications, which transduce signals and most directly affect cell function. Moreover, most drugs target proteins and a better understanding of disease specific changes in the proteome is likely to lead to the development of more, and perhaps better, drugs. Proteomics analyses based on two-dimensional gel analyses and/or mass spectrometry have remained somewhat limited in their application to date. As Jozwik et al. note [11], challenges remain with developing massively parallel proteomic approaches that are adequately sensitive, quantitative, rapid, and robust. These authors propose that one of the two primary protein microarray technologies, where either a series of antibodies (forward capture) or the antigen/specimens (reverse capture) are single spot printed onto slides. While each is limited by the availability of antibodies with adequate sensitivities and specificities, where appropriate antibodies are available these approaches offer significant potential. Of the two approaches, perhaps the reverse capture method has been the more widely adopted. Jozwik et al. provide guidance into the successful application of reverse capture methods, with a focus on studies that use body fluids other than serum/plasma.

The elimination of damaged, improperly folded, or unfolded proteins is a critical homeostatic process in cells. The presence of such proteins can be sensed by the unfolded protein response (UPR), which can then activate the two pathways primarily perform this function – one or more of the autophagy pathways (macroautophagy, microautophagy, chaperone mediated autophagy) [12] and the ubiquitin-proteasome pathway [13]. Generally, the autophagic processes manage proteins with long half-lives, whereas the ubiquitin-proteasome system (UPS) eliminates more short-lived proteins. Targeting autophagy has been used effectively in several experimental cancer models. There is significant cross talk between the UPR, autophagy, and cell death (apoptosis) [14], and inhibiting autophagy reverses resistance to endocrine agents in breast cancer [15, 16]. Mooneyham and Bazzaro [17] provide an excellent assessment of the concepts supporting targeting autophagy and the deubiquitinating enzymes (DUBs) of the UPS as a new approach to eliminating cancer cells. The chapter carefully describes critical cross talk between autophagy and UPS, and addresses the role of autophagy in cancer (both prodeath and pro-survival cell actions).

In their chapter, Li et al. [18] address the need to probe DUB activities experimentally. The authors describe a novel approach to

synthesizing a new class of deubiquitin DUB probes. They show that two ubiquitin moieties, connected by a linker that traps the DUB active site cysteine, offers a new series of reagents to study DUB activities. Detailed materials and methods provide guidance to enable others to use their approach in their own research.

A critical cell survival signaling driven by the UPR is the unconventional splicing of XBP1 following activation of the endogenous RNase activity of IRE1α. For example, in breast cancer cells, activation of XBP1 can drive the ability to grow in the absence of estrogens and resist the antineoplastic effects of antiestrogens [19–21]. IRE1α also has a kinase domain and Feldman and Maly [22] describe biochemical approaches to explore profiling both the kinase and RNase activities of this critical regulator of the UPR. They discuss the role of the soluble form of the protein (IRE1α*), which exhibits many activities seen in the membrane bound form. A greater understanding of the multiple actions of IRE1α, likely will lead to obtaining fundamentally new insights into its role within the UPR and perhaps beyond.

Cell survival signaling, including that regulated by the UPR and autophagy, often includes the regulation of intracellular Ca^{++} concentrations and the action of calcium-binding proteins. Leighton et al. [23] describe a novel tumor suppressor gene, the calcium binding protein ANAX7-GTPase that is also a member of the annexin family. The authors explore the characteristics of the ANAX7-GTPase gene/protein and identify its novel role as a potential driver of the metastatic cascade and also as a biomarker of risk in HER2-negative breast cancers. Signaling cross talk with other critical oncogenic factors, including the ability of ANAX7-GTPase to interact with PTEN and EGFR to regulate PI3K-AKT is also clearly delineated. The chapter further explores the role of ANAX7-GTPase in Ca^{++} metabolism, and how ANAX7-GTPase may be a novel target for gene cancer therapy.

Day et al. [24] provide an insightful discussion of the apparent role of tumor necrosis factor-α-inducible protein 8 (TNFAIP8) in cancer biology, with a specific focus on its role in affecting cell survival and tumor progression. Taking a mechanistic approach, the authors show how knockdown of TNFAIP8 regulates the expression of several critical genes that control proliferation and apoptosis. The multifaceted actions of TNFAIP8 include signaling through multiple oncogenes and signaling through HIF-1α. The authors' detailed and effective assessment of TNFAIP8 action provides a framework for future validation studies. As the critical role of TNFAIP8-centric regulated factors becomes clear in the context of cancer cell survival and disease progression, additional emphasis on their role as biomarkers and targets for drug discovery is likely to emerge.

References

1. Lecca P, Re A (2016) Network-oriented approaches to anticancer drug response. Methods Mol Biol—Cancer Gene Netw 1513:101–115

2. Kim JY, Gatenby R (2016) Quantitative clinical imaging methods for monitoring intratumoral evolution. Methods Mol Biol—Cancer Gene Netw 1513:61–79

3. Verma N, Zhu Z, Hangfu D (2016) CRISPR/Cas-mediated knockin in human pluripotent stem cells. Methods Mol Biol—Cancer Gene Netw 1513:119–139

4. Miller DFB, Yan P, Fang F et al (2016) Complete transcriptome RNA-Seq. Methods Mol Biol—Cancer Gene Netw 1513:141–162

5. Wang J, Ye Z, Huang TH et al (2016) Computational methods and correlation of exon-skipping events with splicing, transcription and epigenetic factors. Methods Mol Biol—Cancer Gene Netw 1513:163–170

6. Zhang W, Lee WY, Zilberberg J (2016) Tissue engineering platforms to replicate the tumor microenvironment of multiple myeloma. Methods Mol Biol—Cancer Gene Netw 1513:171–189

7. Ritchie W (2016) microRNA target prediction. Methods Mol Biol—Cancer Gene Netw 1513:193–199

8. Nguyen C, West GM, Geoghegan KF (2016) Emerging methods in chemoproteomics with relevance to drug discovery. Methods Mol Biol—Cancer Gene Netw 1513:11–21

9. Haznadar M, Mathé E (2016) Experimental and study design considerations in uncovering oncometabolites. Methods Mol Biol—Cancer Gene Netw 1513:37–46

10. Marschall AL, Zhang C, Dübel S (2016) Evaluating the delivery of proteins to the cytosol of mammalian cells. Methods Mol Biol—Cancer Gene Netw (this volume)

11. Jozwik C, Eidelman O, Starr J et al (2016) Validation of biomarker proteins using reverse capture protein microarrays. Methods Mol Biol—Cancer Gene Netw 1513:201–207

12. Clarke R, Cook KL, Hu R et al (2012) Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. Cancer Res 72:1321–1331

13. Higgins R, Gendron JM, Rising L et al (2015) The unfolded protein response triggers site-specific regulatory ubiquitylation of 40S ribosomal proteins. Mol Cell 59:35–49

14. Schwartz-Roberts JL, Cook KL, Chen C et al (2015) Interferon regulatory factor-1 signaling regulates the switch between autophagy and apoptosis to determine breast cancer cell fate. Cancer Res 75:1046–1055

15. Cook KL, Wärri A, Soto-Pantoja DR et al (2014) Chloroquine inhibits autophagy to potentiate antiestrogen responsiveness in ER+ breast cancer. Clin Cancer Res 20:3222–3232

16. Clarke R, Tyson JJ, Dixon JM (2015) Endocrine resistance in breast cancer—an overview and update. Mol Cell Endocrinol (in press)

17. Mooneyham A, Bazzaro M (2016) Targeting deubiquitinating enzymes and autophagy in cancer. Methods Mol Biol—Cancer Gene Netw 1513:49–57

18. Li G, Yuan L, Zhuang Z (2016) Chemical synthesis of activity-based diubiquitin probes. Methods Mol Biol—Cancer Gene Netw 1513:223–232

19. Gomez BP, Riggins R, Shajahan AN et al (2007) Human X-Box binding protein-1 confers both estrogen independence and antiestrogen resistance in breast cancer cell lines. FASEB J 21:4013–4027

20. Davies MP, Barraclough DL, Stewart C et al (2008) Expression and splicing of the unfolded protein response gene XBP-1 are significantly associated with clinical outcome of endocrine-treated breast cancer. Int J Cancer 123:85–88

21. Hu R, Warri A, Jin L et al (2015) NFkappaB signaling is required for XBP1 (U and S) mediated effects on antiestrogen responsiveness and cell fate decisions in breast cancer. Mol Cell Biol 35:379–390

22. Feldman HC, Maly DJ (2016) Profiling the dual enzymatic activities of the serine/threonine kinase IRE1a. Methods Mol Biol—Cancer Gene Netw 1513:233–241

23. Leighton X, Eidelman O, Jozwik C et al (2016) ANXA7-GTPase as tumor suppressor: mechanism and therapeutic opportunities. Methods Mol Biol—Cancer Gene Netw 1513:23–33

24. Day TF, Mewani RR, Starr J et al (2016) Transcriptome and proteome analyses of TNFAIP8 knockdown in cancer cells reveal new insights into molecular determinants of cell survival and tumor progression. Methods Mol Biol—Cancer Gene Netw 1513:83–99

Chapter 2

Emerging Methods in Chemoproteomics with Relevance to Drug Discovery

Chuong Nguyen, Graham M. West, and Kieran F. Geoghegan

Abstract

A powerful interplay exists between the recognition of gene families, sensitive techniques in proteomics, and the interrogation of protein function using chemical probes. The most prominent methods, such as affinity capture, activity-based protein profiling and photoaffinity labeling, are extensively reviewed in the literature. Here we briefly review additional methods developed in the past 15 years. These include "stability proteomics" methods such as proteomically analyzed cellular thermal shift assays and the use of chemical oxidation as a probe of structure, the use of multiple bead-linked kinase inhibitors to analyze inhibitor specificities, and advances in the use of proteolysis-targeting chimeras for selective protein elimination.

Key words Chemoproteomics, CETSA, SPROX, DARTS, LiP-SRM, Proteomics, Kinobeads, PROTAC

1 Introduction

Proteomics has developed spectacularly in its short lifetime at a rate of growth and change rapid enough to unnerve the unwary. In its early days, two-dimensional gels were used to resolve hundreds of proteins into discrete spots in a way that highlighted meaningful variations [1], and digestion followed by Edman sequencing or peptide mass-mapping [2] was used to identify proteins of interest. This approach had the strong attribute that attention went directly to any proteins that changed under test, but its coverage of the proteome was generally low.

Rapid advances in mass spectrometry and the power of the SEQUEST algorithm soon converted the field to methods based on digestion and mass spectrometry, a teenage period that roughly covered the first decade of the twenty-first century [3]. As proteomics reaches its mid-twenties, label-based ratiometric methods or label-free alternatives that increase the depth and scope of comparative analyses form a basis for improving the quantitative rigor of global or broadband proteomic analyses. Meanwhile, specialized methods for targeted analysis have been established for cases in

which quantitation is required, as in the urgent but challenging hunt for protein biomarkers of human disease [4].

Interfacing proteomics with drug discovery has also been challenging. As some foresaw [5], seeking the origins of disease in altered protein expression patterns has not been an especially rewarding endeavor, if the criterion for reward is immediate actionability. Instead, it is applications based on the proteomic algorithm but directed to answering questions about the molecular action of drug-like molecules that are coming to the fore in the pharmaceutical industry.

2 Chemoproteomics: Principal Methods

The literature already offers many authoritative reviews of major methods in chemical biology. Rather than adding to the surfeit, we devote this article to a few specialized methods that have either proved their success or are currently raising hopes of new insights. The common thread running through these techniques is the involvement of a drug or potential drug molecule in the experiment, which is the origin of the term "chemoproteomics." Often, but not always, it denotes an experimental approach composed of two parts, a first one in which affinity for a chemical probe is used to select proteins from a complex mixture, and a second in which the essential methods of proteomics (digestion, mass spectrometry and database searching) are used to identify those proteins. Already, however, this definition is being outdated by expansive new ideas.

Classical methods of this kind have already had real impact. Affinity capture of drug targets using small molecules immobilized on beads or captured after target binding via a biotin "handle" is an example. Handa and colleagues [6] used a refined version of this method employing methacrylate-coated nanobeads to uncloak the mechanism of action of the teratogenic drug thalidomide, rehabilitated now as a therapy for blood cancers. Their result has been incorporated directly into an exciting strategy for targeted protein degradation, which we discuss. Another example is photoaffinity labeling, which increasingly is used to identify specific ligand-binding sites on proteins rather than just to identify targeted proteins [7]. A third is the burgeoning strategy known as activity-based protein profiling [8, 9], which unifies previously disconnected chemical probes that react covalently with protein targets. Although these reagents function by a variety of mechanisms, some assisted by substrate-recognition elements of enzyme active sites and others targeting enzymes purely by complementary reactivity, they are now collectively designated as activity-based probes of enzyme targets. Their defining characteristic is an ability to label covalently a class of enzymes or other proteins that have an important feature in common, such as an exceptionally reactive thiol group or the common capacity to bind ATP.

Instead of providing still more coverage of these well-reported topics, we devote this section to a few technologies that may be new to readers who specialize in fields other than chemical biology or proteomics. These methods are not all new, but we share them in the spirit of colleagues sharing recent reading, as they are gaining momentum and the potential power of the methods reviewed deserves to be appreciated.

3   Stability Proteomics: Cellular Thermal Shift Assay and SPROX

Typical proteomics studies correlate stress on a system (e.g., the presence of a disease state or bioactive compound) with shifts in the expression levels or modifications of specific proteins. Changes in protein abundance or levels of post-translational modification are inferred from the intensity ratios of peptide ions. This type of approach has been used with increasing success for two decades to study protein networks and lay a foundation for systems biology. Despite their power, these techniques are less suited to detecting indirect effects such as changes in protein interactions that leave protein expression levels unchanged. Some recently developed proteomic techniques combine the power of LC-MS with biophysical or biochemical manipulation to monitor the ligand-induced thermodynamic stabilization of proteins, and we refer to these techniques collectively as "Stability Proteomics".

All stability proteomics techniques use or are compatible with bottom-up (i.e., digest-based) mass spectrometric measurements as a readout for the thermodynamic stability of proteins. Experiments often compare results obtained in the absence or presence of a compound with the intent of identifying proteins that become thermodynamically stabilized or destabilized by the compound of interest. An advantage of these methods is that no chemical derivatization or immobilization of the ligand molecule or target proteins is required. In addition, they can all detect on- and off-target protein–drug interactions as well as direct and indirect binding events. Perhaps the most anticipated potential of stability proteomics techniques is the possibility to identify previously unknown off-target (or even unknown on-target) protein–drug interactions.

Methods of this kind detect changes in the thermodynamic stability of proteins using a variety of biophysical and biochemical mechanisms. The cellular thermal shift assay (CETSA) (Fig. 1) was originally developed using a readout based on western blots [10] but has been recently adapted to LC-MS [11]. CETSA relies on the loss of protein solubility in the thermally denatured state to generate thermal denaturation curves. LC-MS approaches generate these melting curves from reporter ion intensities which approximate the abundance of soluble proteins through their peptide surrogates. The experimental protocol for CETSA is attractively simple, but analysis of the data appears challenging.
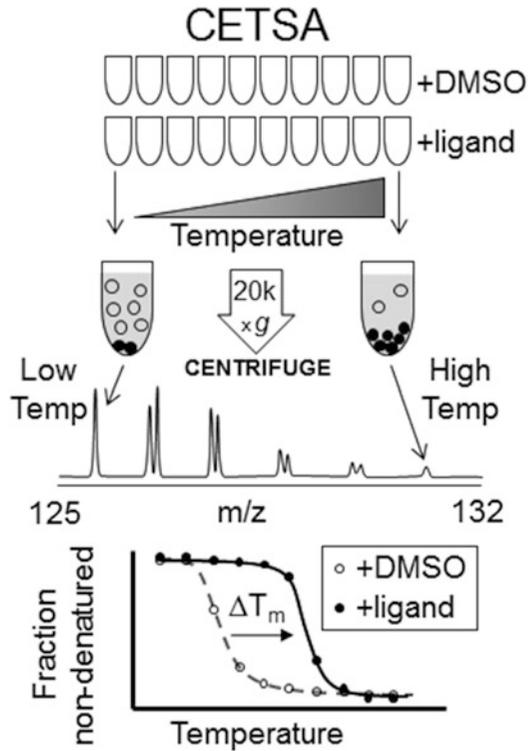
Fig. 1 The CETSA (cellular thermal shift assay) protocol uses ten aliquots of a control (DMSO-treated) sample and ten aliquots of sample treated with ligand. The aliquots can originate from a lysate or intact cell sample. Each aliquot in a set is exposed to one of a series of temperatures and denatured/aggregated proteins are then removed by centrifugation. The fraction of non-denatured protein remaining in each aliquot is measured by TMT (Tandem Mass Tag™) reporter ion intensities from peptides in its tryptic digest, and the reporter ion intensities are used to construct thermal melting curves. Ligand-induced shifts in the T_m values for these curves indicate ligand-induced effects on the thermodynamic stabilities of proteins

An additional method, the "*s*tability of *p*roteins from *r*ates of *ox*idation" (SPROX) technique (Fig. 2) follows the chemical denaturation of proteins by measuring the hydrogen peroxide-mediated oxidation of structurally protected methionines over a range of concentrations of chemical denaturant [12–14]. Unlike the other stability proteomics techniques, SPROX provides additional information regarding affinity (K_d), binding pocket location and domain-specific interactions, but is limited at the peptide level to globally protected methionine-containing peptides.

The approach known as "drug affinity-responsive target stability" (DARTS) [15, 16] and a similar energetics-based method for target identification [17] (Fig. 3) rely on the classical observation that ligand binding often stabilizes proteins against proteolytic degradation. Peptide peak intensities in proteomic analysis are used as the readout for these techniques. DARTS has the
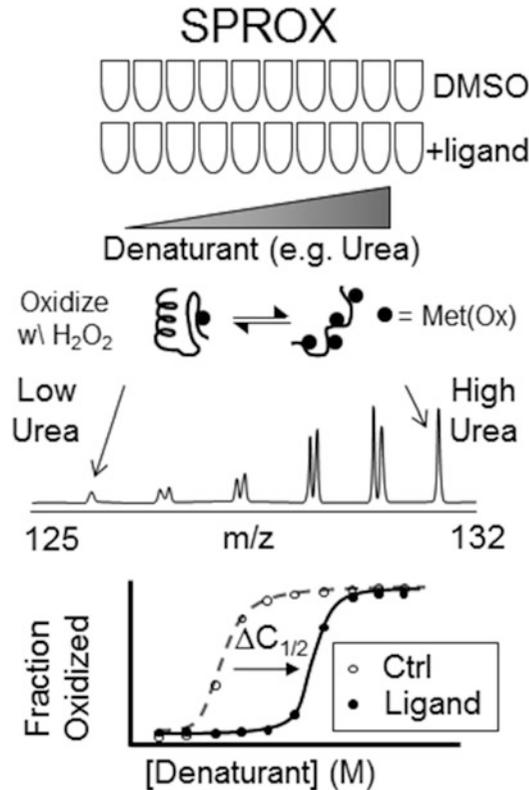
Fig. 2 As with CETSA, the SPROX protocol starts with aliquots of a protein mixture treated with either DMSO (as a control) or ligand, but in this case chemical denaturant is used to shift the folding equilibria of the proteins. Hydrogen peroxide is then added to the aliquots so that solvent–exposed methionine side chains are labeled with oxygen. The extent of methionine oxidation is used to report on the folding equilibrium at each denaturant concentration, typically measured by reporter ions. The data are used to generate denaturation curves, or SPROX curves, and the transition midpoints ($C_{1/2}$) from the SPROX curves for each protein/peptide in the presence and absence of ligand are used to determine changes to proteins' thermodynamic stability in solution at the time of labeling

advantage of being label-free, but the experiment requires prior optimization of conditions.

Limited proteolysis (LiP) coupled with single reaction monitoring (SRM) is similar to DARTS (Fig. 3), but follows the partial digestion step with an additional alternative proteolytic digestion step to increase the number of peptides amenable to bottom up LC-MS and takes advantage of SRM for improved quantitative accuracy. However, LiP-SRM requires additional optimization steps and the quantitative advantage provided by SRM requires prior knowledge of the targets. As such, Lip-SRM may prove to be particularly useful as a follow-up technique for the other stability proteomics techniques that validates potential hits with greater sensitivity.
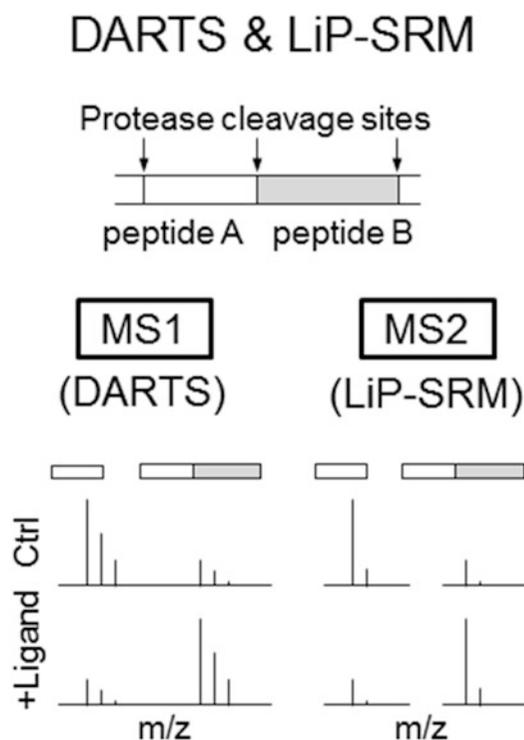
Fig. 3 DARTS (drug affinity-responsive target stability) and LiP-SRM (limited proteolysis coupled with single reaction monitoring) both rely on a protein's susceptibility to protease-catalyzed cleavage. The premise of these techniques is that specific regions of proteins are less susceptible to proteolytic cleavage when they are stabilized by a ligand. Protease efficiency is measured at the peptide level in DARTS and at the peptide fragment level in LiP-SRM. All of the proteomic stability techniques are capable of detecting direct on-target and off-target interactions with ligands (stabilization) as well as indirect effects such as the disruption of a protein–protein interaction (destabilization)

The remainder of this section will focus on CETSA and SPROX, which are the only stability proteomics methods that can generate denaturation curves. Shifts in these curves induced by compound binding can be instructive.

The CETSA technique (Fig. 1) essentially performs a thermal shift assay (TSA) on every protein in a given test mixture. As with TSA of a pure protein, in which the protein is carried through its denaturation equilibrium over a series of increasing temperatures, the CETSA technique gauges the relative amount of each protein that remains folded over a range of temperatures. Unlike most TSA techniques for single pure proteins, CETSA estimates denaturation based on solubility rather than fluorescence. The CETSA derived "T_m" value may not be accurate for proteins that unfold irreversibly, but this does not prevent the technique from detecting ligand-induced shifts in CETSA curves. The fact that CETSA can

screen for ligand interactions from (potentially) every peptide/protein detected in an LC-MS experiment makes this new technique appear very appealing. In addition, it is reported to be able to be performed on live cells as well as on cell lysates.

The process by which denaturation curves are generated involves dividing a protein mixture or cell population into equal aliquots. Each aliquot is heated to a different temperature that will contribute a point to the thermal melting curve for each protein detected. The temperature-specific samples are lysed, if necessary, and centrifuged to remove proteins which are no longer soluble as a result of denaturation. After digestion, isobaric tagging and pooling, the sample is analyzed by LC-MS and thermal denaturation curves are generated for each detected peptide from the isobaric tag reporter ions. These peptide-specific data must then be pooled to give data at the protein level.

An additional sample containing a compound of interest is prepared likewise and the CETSA denaturation curves are compared between conditions to identify proteins that were stabilized or destabilized by the compound. In addition, CETSA can also generate isothermal dose–response (ITDR) profiles that resemble dose–response curves.

The SPROX technique provides results fundamentally similar to chemical denaturation curves from CD or fluorescence-based studies on pure proteins [14, 18]. SPROX methodology calls for diluting a protein mixture into a series of chemical denaturants that shift the unfolding equilibria of proteins in each denaturant-containing mixture (Fig. 2). Solvent-accessible Met side chains are then labeled (oxidized) with hydrogen peroxide. The oxidant is quenched, and the protein samples are digested, labeled with isobaric tags and pooled. SPROX uses the denaturant-dependent oxidation of Met side chains to generate denaturation curves and measure thermodynamic properties of proteins. Reporter ion intensities from oxidized and unmodified methionine-containing peptides are used to generate this data and the folding free energy (ΔG), m-value, and K_d can be derived by fitting this data to SPROX equations. However, these ΔG values may not be accurate for cases where oxidation affects protein equilibria. Although the K_d values may still be accurate when ΔG estimation is compromised, the accuracy of K_d values is affected when oxidation interferes with compound binding. The number of proteins identified from an LC-MS experiment that can be useful in a SPROX analysis is also limited to those proteins for which digestion yields Met-containing peptides. In theory the SPROX approach could generate dose–response data similar to PLIMSTEX (protein–ligand interaction by mass spectrometry) [19], but this has yet to be demonstrated.

As with all proteomic studies, novel hits from stability proteomics techniques require validation using orthogonal techniques such as western blotting with immunodetection. Nonetheless, the

various emerging methods of stability proteomics present researchers with the exciting potential to discover novel on- and off-targets of their compounds. Having knowledge of all targets for a given drug candidate could enable medicinal chemists to design around off-target interactions that could be potentially detrimental in vivo and in the clinic. On a more cautionary note, the widespread feasibility and applicability of these techniques remains to be seen, as the few novel target identifications secured by these methods to date all come from the labs in which the techniques were developed. Although the SUPREX technique (stability of unpurified proteins from rates of H/D exchange), on which SPROX was based, has seen broader use among the scientific community [20, 21], time will tell if these new stability proteomics techniques will be adopted into the tool kits of external laboratories.

4   Kinobeads

The central importance of protein kinases in biology places them among the most attractive drug targets, but the extent of their mutual similarity initially made specific inhibition of particular kinases appear to be a difficult task [22]. Persistence has transformed the picture and kinase inhibitors now appear in the clinic in the form of a variety of life-saving and life-enhancing medicines. In some cases, they are less specific than was originally intended, and this apparent shortcoming has turned out to be a strength [23]. Therefore, the ability to define in quantitative terms the specificity of protein kinase inhibitors is a topic of great importance.

Kinase selectivity screens using panels of enzymes [24] and a notably successful ATP-based covalent probe [25] have provided two routes to acquiring the needed information, but one of the most streamlined and elegant approaches is the use of kinobeads [26] (Fig. 4). This approach exploits the ability of multiple kinase inhibitors linked to agarose beads to capture a high fraction of the protein kinases present in a cell or tissue lysate. Inhibitors present in soluble form can compete against the bead-linked compounds, allowing a direct means of gauging the relative affinities of different enzymes and inhibitors.

The kinobead approach has some attractive features. Coupling kinase inhibitors that belong to several different classes to agarose beads has allowed the capture and MS-based identification of as many as several hundred kinases per assay, while the use of isobaric chemical tags (initially iTRAQ) [27] allows for ratiometric quantitation of their binding or displacement by competitors. Varying the concentrations of soluble inhibitor during the capture step allows derivation of an IC50 for any detected kinase.

To introduce the concept and its utility, Bantscheff and coworkers investigated the inhibitory activities of several drugs, imatinib
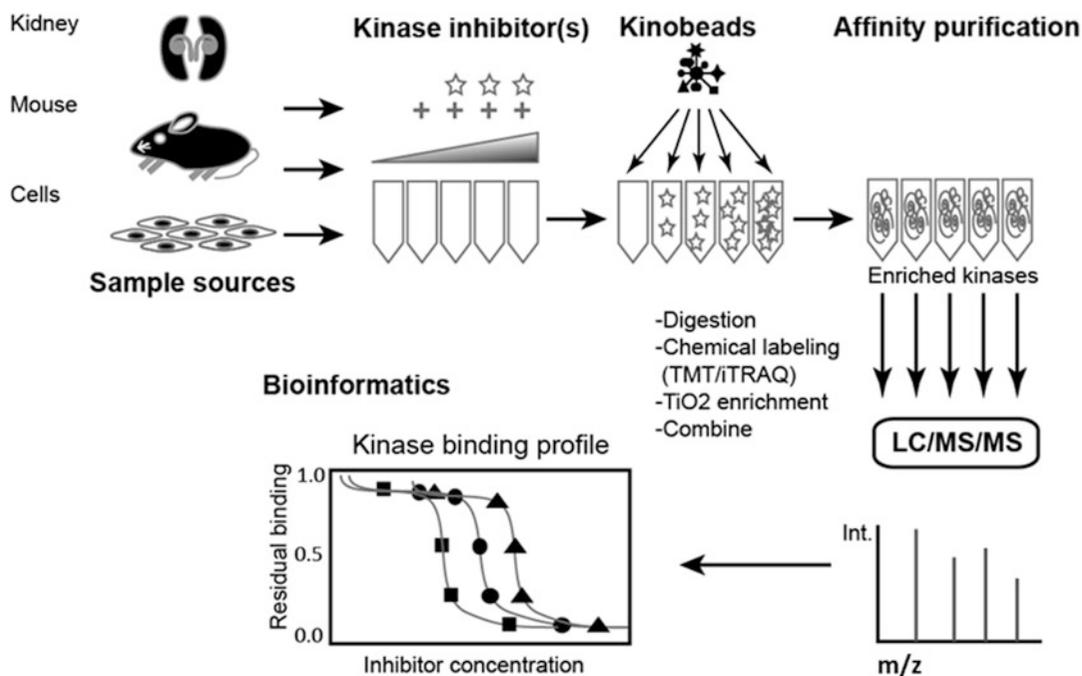
Fig. 4 Kinobeads workflow. Kinase-containing samples are first incubated with increasing concentrations of a kinase inhibitor. Equal amounts of kinobeads (polymeric beads carrying multiple immobilized kinase inhibitors) are then added to each sample, leading to capture of protein kinase molecules while inhibitor-bound kinase molecules remain in solution. The beads are washed, the proteins bound to them are digested, and the multiple digests are first chemically labeled with isobaric tagging reagents such as TMT or iTRAQ, and then combined for ratiometric quantitation of the kinases bound to the beads and the effect on their binding of the soluble competitor. Label-free quantitation (LFQ) might be used as an alternative, in which case no mixing would occur. Kinase binding as a function of inhibitor concentration is extracted from the result, and the respective affinities of the soluble inhibitor for different kinases can be calculated

(Gleevec), dasatinib (Sprycel), and bosutinib (Bosulif) [26]. Lysates deriving from K562 cells, which express BCR-ABL fusion protein, were first incubated (separately) with an increasing concentration of each kinase inhibitor (100 pM–10 μM), after which the lysates were exposed to kinobeads. Kinases captured by the beads despite the competitive presence of the soluble inhibitor were digested, labeled with iTRAQ reagent, and detected by mass spectrometry with reporter ion readouts providing a gauge of the extent of binding at each concentration of the competitor. This allowed IC50 values for all three drugs to be derived.

An additional benefit of the work was its potential to identify nonkinase drug targets. In the introductory paper [26], potent inhibition of the oxidoreductase NQO2 by imatinib was indicated by the kinobeads binding protocol and directly confirmed in an enzyme assay.

Several improvements to the workflow and new applications for the method have now been reported [28, 29]. The original

format held the possibility that endogenous ATP or related compounds might compete with soluble kinase inhibitors for occupancy of ATP binding sites on kinases during the first step, potentially causing IC50 values to be inconsistent between experiments. To correct for this, lysates were first depleted for cofactors by gel filtration before addition of inhibitors. As expected, the concentrations of endogenous nucleotide factors can greatly affect the observed affinities of inhibitors for target kinases [29].

Although the kinobeads method was devised to profile kinase inhibition, it has been adapted to measuring differences in the expression of kinases between two cell lines. To do this, kinobeads were used to capture kinases from different cell lines without any resort to competition from soluble inhibitors. The relative expressions of kinases were derived by taking the summation of their corresponding three largest peptide intensities [29, 30].

To broaden the scope of the kinobead method, Médard et al. synthesized a new generation of kinobeads capable of capturing a wider selection of kinase families [28]. The workflow was streamlined by performing competition assays in 96-well plates and label-free quantitation in the powerful MaxQuant program [31] was used instead of iTRAQ. The improved method allowed 216 protein kinases to be captured and, therefore, potentially to be the targets of competitive binding studies.

Finally, comparison of the kinobeads method with covalent capture of kinases using acylphosphate ATP analogs indicated that the two methods are complementary [32] and can be used in tandem when maximum coverage of the kinase complement of a sample is required.

5   Targeted Protein Degradation

As we noted above, affinity-based protein capture allied to sensitive protein identification was the method that resolved the molecular mechanism of thalidomide [6]. The drug and its relatives bind to cereblon, the substrate recognition module of a certain E3 ubiquitin protein ligase complex (there are more than 600 [33]), and can modulate its specificity for certain protein substrates. For example, exquisite SILAC-based proteomic analysis was used to demonstrate the extraordinarily specific effect of lenalidomide in bringing about the ubiquitination and degradation of casein kinase 1α as well as two transcription factors, IKZF1 and IKZF3 [34]. The latter effects provide the clinical efficacy of lenalidomide in multiple myeloma.

Increased understanding of the mode of action of these compounds has led to their incorporation into the existing strategy of using a double-headed drug to bring a protein targeted for elimination into close proximity to an E3 ubiquitin ligase. One end of the agent should have affinity for the targeted protein, and has been a drug-like small molecule from the outset. The element to

be recognized by the E3 ubiquitin ligase was originally a peptide [35], but difficulty with cell-permeability caused this to give way to an E3-targeting small molecule [36, 37]. Impressive preliminary demonstrations of the use of a thalidomide-related E3-targeting group have appeared [38, 39], and interest in this strategy is sure to continue to grow despite concerns that the method requires agents that embody two drug-like moieties and consequently will tend to have molecular weights beyond the preferred range for good pharmacokinetic properties.

6   Conclusion

The power and scope of rapidly emerging methods in chemical biology is one of many developments resulting from the genomic revolution that occurred mainly in the 1990s. This resulted in massive enhancement of our understanding that related proteins are derived from families of related genes. Together with the emergence of methods for protein recognition based on the essential algorithm of proteomics, the stage has been set for adventurous exploration of the potential of chemical probes—agents that address proteins with respect to their functions—to elucidate details of the subtle differences that may exist between related family members. As proteins will presumably continue to account for the great majority of drug targets, this additional capability to monitor their functional aspects will be an important complement to mainstream drug discovery in the coming decades.

References

1. O'Farrell PH (2008) The pre-omics era: the early days of two-dimensional gels. Proteomics 8:4842–4852

2. Henzel W, Watanabe C, Stults J (2003) Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom 14:931–942

3. Patterson SD, Aebersold RH (2003) Proteomics: the first decade and beyond. Nat Genet 33:311–323

4. Gillette MA, Carr SA (2013) Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. Nat Methods 10:28–34

5. Cohen J (2001) The proteomics payoff. Technol Rev 104:54–60

6. Ito T, Ando H, Suzuki T et al (2010) Identification of a primary target of thalidomide teratogenicity. Science 327:1345–1350

7. Geoghegan KF, Johnson DS (2010) Chemical proteomic technologies for drug target identification. Annu Rep Med Chem 45:345–360

8. Niphakis MJ, Cravatt BF (2014) Enzyme inhibitor discovery by activity-based protein profiling. Annu Rev Biochem 83:341–377

9. Cravatt BF, Wright AT, Kozarich JW (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu Rev Biochem 77:383–414

10. Molina DM, Jafari R, Ignatushchenko M et al (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science 341:84–87

11. Savitski MM, Reinhard FBM, Franken H et al (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. Science 346:55

12. Adhikari J, West GM, Fitzgerald MC (2015) Global analysis of protein folding thermodynamics for disease state characterization. J Proteome Res 14:2287–2297

13. Strickland EC, Geer MA, Tran DT et al (2013) Thermodynamic analysis of protein-ligand binding interactions in complex biological

mixtures using the stability of proteins from rates of oxidation. Nat Protoc 8:148–161

14. West GM, Tang L, Fitzgerald MC (2008) Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass spectrometry-based strategy. Anal Chem 80:4175–4185

15. Lomenick B, Olsen RW, Huang J (2011) Identification of direct protein targets of small molecules. ACS Chem Biol 6:34–46

16. Lomenick B, Hao R, Jonai N et al (2009) Target identification using drug affinity responsive target stability (DARTS). Proc Natl Acad Sci U S A 106:21984–21989

17. Liu PF, Kihara D, Park C (2011) Energetics-based discovery of protein-ligand interactions on a proteomic scale. J Mol Biol 408:147–162

18. West GM, Tucker CL, Xu T et al (2010) Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. Proc Natl Acad Sci U S A 107:9078–9082

19. Zhu MM, Rempel DL, Du Z, Gross ML (2003) Quantification of protein-ligand interactions by mass spectrometry, titration, and H/D exchange: PLIMSTEX. J Am Chem Soc 125:5252–5253

20. Tang L, Sundaram S, Zhang J et al (2013) Conformational characterization of the charge variants of a human IgG1 monoclonal antibody using H/D exchange mass spectrometry. MAbs 5:114–125

21. Liyanage R, Devarapalli N, Puckett LM (2009) Comparison of two ESI MS based H/D exchange methods for extracting protein folding energies. Int J Mass Spectrom 287: 96–104

22. Cohen P (2002) Protein kinases—the major drug targets of the twenty-first century? Nat Rev Drug Discov 1:309–315

23. Faivre S, Demetri G, Sargent W, Raymond E (2007) Molecular basis for sunitinib efficacy and future clinical development. Nat Rev Drug Discov 6:734–745

24. Karaman MW, Herrgard S, Treiber DK et al (2008) A quantitative analysis of kinase inhibitor selectivity. Nat Biotechnol 26:127–132

25. Patricelli MP, Szardenings AK, Liyanage M et al (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry 46:350–358

26. Bantscheff M, Eberhard D, Abraham Y et al (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nat Biotechnol 25:1035–1044

27. Ross PL, Huang YN, Marchese JN et al (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154–1169

28. Médard G, Pachl F, Ruprecht B et al (2015) Optimized chemical proteomics assay for kinase inhibitor profiling. J Proteome Res 14:1574–1586

29. Becher I, Savitski MM, Savitski MF et al (2013) Affinity profiling of the cellular kinome for the nucleotide cofactors ATP, ADP, and GTP. ACS Chem Biol 8:599–607

30. Grossmann J, Roschitzki B, Panse C et al (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 73:1740–1746

31. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372

32. Lemeer S, Zörgiebel C, Ruprecht B et al (2013) Comparing immobilized kinase inhibitors and covalent ATP probes for proteomic profiling of kinase expression and drug selectivity. J Proteome Res 12:1723–1731

33. Cohen P, Tcherpakov M (2010) Will the ubiquitin system furnish as many drug targets as protein kinases? Cell 143:686–693

34. Krönke J, Fink EC, Hollenbach PW et al (2015) Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. Nature 523:183–188

35. Zhou P, Bogacki R, McReynolds L, Howley PM (2000) Harnessing the ubiquitination machinery to target the degradation of specific cellular proteins. Mol Cell 6:751–756

36. Carmony KC, Kim KB (2012) PROTAC-induced proteolytic targeting. Methods Mol Biol 832:627–638

37. Sakamoto KM (2010) Protacs for treatment of cancer. Pediatr Res 67:505–508

38. Bondeson DP, Mares A, Smith IED et al (2015) Catalytic in vivo protein knockdown by small-molecule PROTACs. Nat Chem Biol 11:611–617

39. Lu J, Qian Y, Altieri M et al (2015) Hijacking the E3 ubiquitin ligase cereblon to efficiently target BRD4. Chem Biol 22:755–763

Chapter 3

ANXA7-GTPase as Tumor Suppressor: Mechanisms and Therapeutic Opportunities

Ximena Leighton, Ofer Eidelman, Catherine Jozwik, Harvey B. Pollard, and Meera Srivastava

Abstract

Chromosomal abnormalities, including homozygous deletions and loss of heterozygosity at 10q, are commonly observed in most human tumors, including prostate, breast, and kidney cancers. The ANXA7-GTPase is a tumor suppressor, which is frequently inactivated by genomic alterations at 10q21. In the last few years, considerable amounts of data have accumulated describing inactivation of ANXA7-GTPase in a variety of human malignancies and demonstrating the tumor suppressor potential of ANXA7-GTPase. ANXA7-GTPase contains a calcium binding domain that classifies it as a member of the annexin family. The cancer-specific expression of ANXA7-GTPase, coupled with its importance in regulating cell death, cell motility, and invasion, makes it a useful diagnostic marker of cancer and a potential target for cancer treatment. Recently, emerging evidence suggests that ANXA7-GTPase is a critical factor associated with the metastatic state of several cancers and can be used as a risk biomarker for HER2 negative breast cancer patients. Cross talk between *ANXA7*, *PTEN*, and *EGFR* leads to constitutive activation of PI3K-AKT signaling, a central pathway of tumor cell survival and proliferation. This review focuses on the recent progress in understanding the tumor suppressor functions of ANXA7-GTPase emphasizing the role of this gene in Ca^{2+} metabolism, and exploring opportunities for function as an example of a calcium binding GTPase acting as a tumor suppressor and opportunities for ANXA7-GTPase gene cancer therapy.

Key words ANXA7-GTPase, Tumor suppressor gene, Cancer, Calcium, Apoptosis

1 Introduction

The rising incidence of cancer and coincident morbidity and mortality are significant health problems and cost enormous health resources. Despite recent intensive research investigations, much remains to be learned about specific molecular defects associated with onset and progression of cancer. Current evidence suggests that the development of cancer may be driven, at least in part, by

the inactivation of as yet unknown tumor suppressor genes (TSG) which are presently identified only as sites of allelic loss. The identification and characterization of these TSGs are useful for establishing molecular markers which can identify patients at risk for advanced disease. TSGs function properly in normal tissues by regulating the growth of normal cells. Classically, TSGs have been suspected on the basis of noting a specific relationship between the occurrence of tumor(s) and chromosomal defects. However, mutations, deletions, or other modes of inactivation of these TSGs should also be expected to contribute to uncontrolled growth and malignant transformation of normal cells. Many TSGs have been cloned from humans and found to be mutated in a variety of human cancers. In this review, we discuss a new TSG, ANXA7-GTPase, which differs from other TSGs in terms of its unique chemical, biophysical, and biological properties.

2   Identification of *ANXA7* Gene on Chromosome 10q21

Multiple potential tumor suppressor genes have been hypothesized to exist around the 10q21 locus of the chromosome 10. Examples of the diseases correlated with loci at chromosome 10 include myxoid chondrosarcoma (10q21.1, [1]); sporadic nonmedullary thyroid carcinoma (10q21.1, [2]); renal cell carcinoma (10q21-23, [3]); chronic myelogenous leukemia (10q21, [4]); glioma (10q21-26, [5]); glioblastoma (two independent regions: 10pter-q11 and 10q24-q26, [6]); colonic adenocarcinoma (inverted, non-ret duplication of 10q11 to 10q21, [7]); lung carcinoma (10q21-10qter, [8]); hepatocellular carcinoma (10q, [9]); and prostate cancer (two independent loci:10q21 and 10q23-24, [10]). A frequently deleted locus on chromosome 10q24-25 has recently been shown to harbor the PTEN tumor suppressor gene [11, 12], thus supporting the concept of multiple candidate tumor suppressor genes in this region. The *ANXA7* gene is located at human chromosome 10q21, and *ANXA7* deletions are extremely common in human cancers, especially those that arise from prostate and breast. Four microsatellite markers at or near the *ANXA7* locus on chromosome 10q21 were used to analyze the DNA from laser capture microdissected tumor cells and matched normal cells. Thirty-five percent of the 20 prostate tumors showed loss of heterozygosity (LOH). The microsatellite marker closest to the *ANXA7* locus showed the highest rate of LOH, including one homozygous deletion [13]. Similarly, 40 % of the informative breast cancer patients were found to have LOH [14]. The finding of a tumor suppressor gene *ANXA7* in this chromosomal region with frequent mutations/deletions in human prostate cancers raises important questions as to its potential contribution to this cancer type.

3 Expression of ANXA7 in Cancer

Annexins are a family of calcium-dependent, phospholipid and membrane binding proteins containing a conserved repeating domain of approximately 70 amino acids [15–17]. These are thought to be associated with membrane trafficking, signal transduction, cellular differentiation and proliferation [16, 18–21]. Annexins have also been shown to function as ion channels [21–23]. Dysregulated expression of several members of the annexin family of proteins has been observed in prostate cancer [13, 24–27]. The expression of ANXA1 (lipocortin I) and ANXA2 (calpactin I) has been reported to be decreased in androgen-stimulated prostate cancer compared with benign prostatic epithelium [23, 24, 26]. Conversely others have noted no ANXA1 or ANXA2 protein expression changes in androgen-stimulated prostate cancer, while reporting a significant decrease in recurrent prostate cancer expression [27]. The expression of ANXA7 (synexin) has been shown to be commonly lost in hormone-refractory and locally recurrent metastatic prostate cancer [13, 27, 28], glioblastoma [29, 30], and melanoma [31]. Numerous examples of ANXA7 protein distribution in different tumors unequivocally confirmed its tumor suppressor effect which can be plausibly associated with the regulation of immune response and hormonal balance through other ANXA7-mediated major effects—the Ca(2+)/GTP-dependent secretion and exocytosis.

One of the most significant features of ANXA7 is its differential distribution in cancer compared with equivalent normal tissues. Early work on the *ANXA7* gene has shown that it is expressed in small amounts in nearly every normal cell, and that it is found throughout phylogeny as a single copy gene in organisms as diverse as man [32], mouse [33, 34], Xenopus [35], and Dictyostelium [36, 37]. Two types of molecular abnormalities have been reported that might contribute to aberrant ANXA7 expression at least in certain cancers. Dramatic overexpression of ANXA7 compared with normal tissues has been demonstrated in tumors of breast, liver, nasopharyngeal, colorectal, testis, thyroid, and oral cavity [38–43]. Conversely, loss of ANXA7 expression has been demonstrated in prostate, skin, glioblastoma, melanoma, pancreas and adrenal gland neoplasms [13, 29–31, 42, 43]. A comprehensive survey of ANXA7 protein expression in patient-matched benign prostatic epithelium, high grade prostatic intraepithelia neoplasms (PIN), stage T2 and T3/4 primary tumors, metastatic, and locally recurrent prostate cancer has revealed that significant reductions in ANXA7 expression occur in a stage specific manner [13]. ANXA7 expression is completely lost in a high proportion of metastases (57%) and in local recurrences of hormone refractory prostate cancer (63%). The significance of the difference from other

samples is at the $p = 0.0001$ level. On the other hand, the expression of ANXA7 in breast cancer is significantly enhanced and has been associated with the presence of metastatic disease ($p < 0.0001$) [44, 45]. These studies suggest that the dysregulations of ANXA7 are important events in prostate and breast carcinogenesis and cancer progression.

These data illuminate a profound difference in the biological consequences of ANXA7 expression in breast and prostate cancers. A similar disparity has previously been found for the antiapoptotic protein Bcl-2. High Bcl-2 predicts adverse prognosis in primary prostate cancer [46], but is a favorable prognostic marker in breast cancer [47, 48]. Bcl-2 expression in normal prostate glands is restricted to the basal cell compartment [46], whereas ANXA7 is preferentially expressed in the secretory cells of the prostate [13]. Thus, reappearance of basal cell like gene expression patterns in a fraction of cancers may be interpreted as a stem cell effect leading to dedifferentiation and more aggressive behavior. While prostatic stem cells are located in the basal cell layer [49], the exact identity of the putative stem cells of the mammary gland remains a matter of investigation [50, 51]. Interestingly, ANXA7 has pro-apoptotic properties as compared to the apoptosis inhibiting function of Bcl-2. The opposite role in apoptosis control of Bcl-2 and ANXA7 may explain their opposite expression pattern in benign prostate glands and prostate cancers (Table 1). In addition, loss of ANXA7 or overexpression of Bcl-2 leads to downregulation of IP3 Receptor-regulated store-operated calcium channels. We conclude that the reason for the different ANXA7 expression patterns between breast and prostate cancer may be explained by differential regulation and functions of ANXA7, and the relative importance of pro-and anti-apoptotic and calcium signaling pathways in these tissues. It is plausible that multiple signaling pathways might converge on the regulation of the *ANXA7* gene in neoplasia.

Table 1
Biological basis for differences in ANX7 expression in breast and prostate cancers

	Prostate		Breast	
	Bcl-2	ANXA7	Bcl-2	ANXA7
Normal tissue	↓	↑	↑	↓
Stem cell	↑	↓	?	?
Cancer	↑	↓	↓	↑
Favorable prognosis	↓	↑	↑	↓

4   ANXA7 Knockout Mouse

Reexpression of ANXA7 protein in human tumor cells kills cancer cells by apoptosis and causes tumor suppression in nude mice. Although the early kinetics of ANXA7 inactivation is carcinogenesis and the human studies support assignment of ANXA7 as a tumor suppressor, many cancer geneticists consider the outcome in the gene knockout mouse model as the ultimate biological test. If ANXA7 loss in humans does contribute to cell transformation, then murine ANXA7-deficient animals might be expected to develop tumors, potentially in the same organs in which human ANXA7 loss-initiated tumors appear. This was precisely the result in the Anxa7 knockout mice, all the more striking because this was after genetically inactivating only one *Anxa7* allele in the mouse. Because the murine *Anxa7* locus, like its syntenic human counterpart, is fragile, mice that were heterozygous for inactivation of *ANXA7* developed mostly lymphosarcoma of the thymus, spontaneously due to genomic instability. In general, the frequency of spontaneous tumor occurrence is in the range of 20–30 %, becoming more accentuated with advancing age [52]. Importantly the homozygous knockout (−/−) mouse has a lethal phenotype, meaning that the *Anxa7* gene is crucial for life and has no genetic "backups."

Increased tumor incidence in the *Anxa7*-challenged mice was associated with the gender-related cell growth anomalies indicating possible defects in sex hormone synthesis or actions. While only male *Anxa7* (+/−) mice had extraordinarily enhanced growth spurt including internal organs, the female mice had significantly higher incidence of tumors compared to males. Some gender-related anomalies in major lymphatic organs possibly reflected defective immune response in tumorigenesis: unlike splenomegaly present in males, spleen size was decreased in females, while lymphosarcoma (thymus, spleen) was the most frequent tumor, especially in females. Along with growth anomalies and susceptibility to tumorigenesis, the *Anxa7*-challenged mice displayed multiple defects in their endocrine status. In the *Anxa7* (+/−) mice the genes responsive to the nutritional state were set at constantly higher levels of expression resulting in the phenotype of adrenal medullary hypertrophy and chromaffin cell hyperplasia, which can be related to chromogranin A, a master control "switch" for dense core granule formation [53–55]. Islets of Langerhans in these mice exhibited a profound reduction in ITPR3 protein expression, defective intracellular calcium signaling as well as insulin secretion which presumably contributed to the loss of genomic discrimination between the fed and fasted states [56].

5 ANXA7 and Calcium Homeostasis

The phenotype of *Anxa7* heterozygous mice not only demonstrates that loss of *Anxa7* predisposes to carcinogenesis, but also suggests that an intact calcium homeostasis might either protect the *ANXA7* locus or maintain ANXA7 expression in humans. ANXA7 is a Ca^{2+}-activated GTPase protein with membrane fusion properties which has been implicated in the regulation of exocytotic secretory processes in pancreatic islets and elsewhere. While the nullizygous knockout *Anxa7* gene is lethal, the heterozygous *Anxa7* (+/−) mice survives and expresses an instructive beta cell defect. In studies with β-cells from the viable heterozygous *Anxa7* (+/−) mice, we learned that reduced *Anxa7* gene dosage in the *Anxa7* heterozygous mice not only leads to an aberrant Ca^{2+} effect on insulin secretion, a selective decrease in IP_3-Receptor expression and IP_3-Receptor mediated calcium stores, but also a defective store-operated calcium channel (SOC) [53]. Levels of serum testosterone, estrogen, growth hormone, glucagon, corticosterone, and IGF-1 in the mutants were identical to levels found in the normal littermates, suggesting that the problems were likely to be local to the β-cells [57]. In addition to these defects in Ca^{2+} metabolism in cultured beta cells, we noted that the IP3-mediated calcium transients are attenuated in T-lymphocytes and cerebellum from *Anxa7* (+/−) mice heterozygotes (unpublished data).

6 ANXA7 Gene Mediates Pro-apoptotic Actions of Ca^{2+} in Tumor Cells

ANXA7 has been long known to be a mediator of calcium action, and cells from the cancer-prone *Anxa7* (+/−) knockout mice are also known to be deficient in IP3 Receptors. Since Ca^{2+} is a central actor in the apoptosis process, and cells with loss of IP3 receptors are apoptosis-resistant, we hypothesized that this calcium-focused mechanism may be the basis of ANXA7 tumor suppressor gene activity. If this is true, this will be the first instance in which the activity of a tumor suppressor gene is mediated by Ca^{2+}. Most prostate cancer cells actually divide slowly, and their cancer phenotype therefore appears to be ruled more by inhibition of apoptosis than by enhanced proliferation [58–60]. Calcium elevation is a necessary preliminary event in the initiation of the apoptotic cascade [58, 61, 62], and compounds that recruit intracellular calcium from the endoplasmic reticulum have therefore been increasingly studied as prototype drugs for induction of apoptosis in prostate cancer cells. An example of such a compound is thapsigargin, which arrests prostate cancer cells in G1/G0 of the cell cycle and induces

the apoptotic cascade [58, 63, 64]. Thapsigargin raises cytosolic Ca^{2+} concentration by blocking the SERCA-pump ATPase on the ER, thereby blocking reuptake of calcium into the ER. As a further consequence of elevation of cytosolic free calcium, either by thapsigargin or endogenous mechanisms, plasma membrane-localized store-operated calcium (SOC) channels are also activated. The latter process contributes to elevation of intracellular calcium, and thus further pushes the cell into the apoptotic cascade. As an evidence for the contribution of the SOC channels to the apoptotic signal, Wertz and Dixit [60] have shown that LNCaP cells are quite resistant to pro-apoptotic stimuli if Ca^{2+} release is blocked and SOC channels are silent. The importance of increased cytosolic Ca^{2+} for growth arrest at G1/G0 and induction of apoptosis in prostate cancer cells is further manifested by the fact that the thapsigargin effect can be blocked by the intracellular calcium chelator BAPTA [63], and can be mimicked by addition of the general calcium ionophore ionomycin [64].

Large molecules such as proteins have been shown to have effects similar to those induced by thapsigargin. For example, exogenously added TGF-β also arrests androgen-independent prostate cancer cells in G1/G0, and induces apoptosis [65]. Transfection of the *ANXA7* gene into the DU145 cells also has a very similar sequence of actions. This fact, plus the high prevalence of tumors in the *Anxa7*(+/−) mice and disorders of calcium metabolism in *Anxa7* (+/−) mice tissues appear to implicate a specifically thapsigargin-like mechanism for how the *ANXA7* gene activates human tumor cell apoptosis. For example, in cancer cells, thapsigargin raises cytosolic calcium concentration by preventing reentry of the Ca^{2+} into the ER and activates SOC channels. Similar studies of Ca^{2+} metabolism in beta cells from *Anxa7* (+/−) mice, showed that thapsigargin failed to raise cytosolic Ca^{2+}, and failed to activate SOC channels The ligand IP3 also failed to release intracellular Ca^{2+} from the ER. So the reason for the lack of efficacy of thapsigargin in the *Anxa7* (+/−) mice is a documented tenfold deficiency in IP3 Receptors [53]. The importance of IP3 receptors in cancer cells is that IP3 Receptor activation by IP3 is the physiological stimulus needed to release calcium from the ER, thus triggering the mitochondrial permeability transition, which leads to apoptosis [66]. Therefore, it is possible that the action of the transfected *ANXA7* gene on prostate tumor cells, may be to elevate cytosolic Ca^{2+}, and to potentiate subsequent pro-apoptotic actions of the released calcium. The data supporting this perspective and these actions are summarized in Fig. 1. We presume that any of these steps could be part of the mechanism by which addition of the *ANXA7* gene might induce prostate tumor cell death by apoptosis.
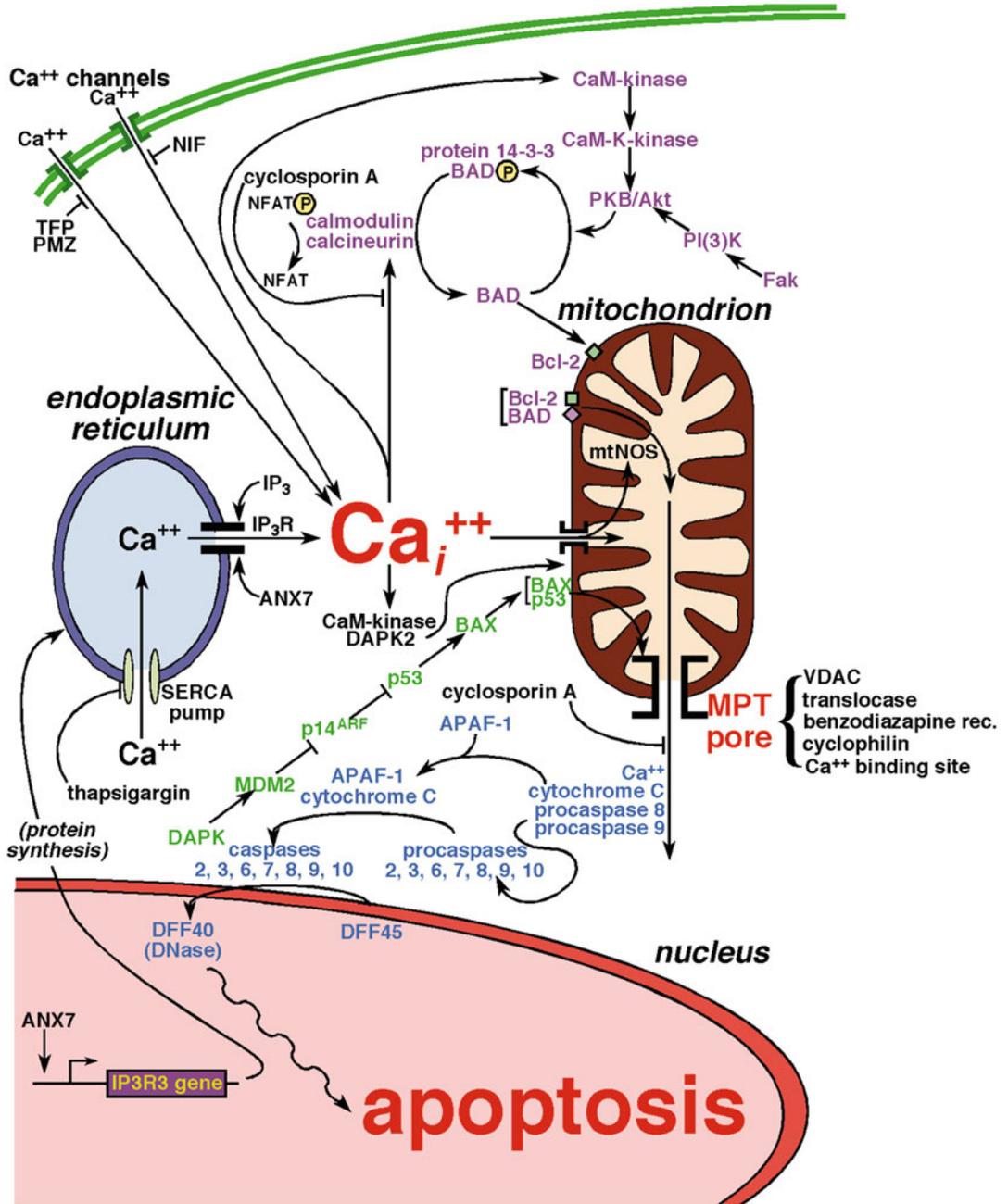
Fig. 1 Schematics of the role of ANXA7 and calcium in apoptosis

7   ANXA7 and Molecular Diagnosis of Cancer

When analyzed retrospectively, cancer patients expressing altered ANXA7 exhibited shortened survival [44, 45], association with unfavorable markers of disease progression, metastasis, and accelerated rates of recurrences. The results obtained from 1077 breast tissue specimens showed that increased ANXA7 expression is associated with metastatic disease and significantly decreased survival in those breast cancer patients who present with BRE grade 2 tumors, or tumors lacking detectable HER2 expression. Cox regression analysis reveals that HER2 negative patients suffer a doubling in the risk of death with each increasing level of ANXA7 expression. Remarkably, in HER2 negative patients the difference in risk is tenfold between those with negative ANXA7 expression, and those with strong ANXA7 expression. The clinical treatment of primary breast cancers has been greatly complicated by the inability to accurately predict which tumors will eventually become invasive and metastatic, and which will become localized and indolent. Strong expression of HER2 in 20–35 % of the breast cancer patients is known to be associated with poor prognosis, and has been used to predict response to treatment with the anti-HER2 antibody trastuzumab (Herceptin®). Recent data suggest that the expression level of ANXA7 can help to stratify the remaining HER2 negative patients who need the most focused attention. At a minimum, the value of these results for HER2 negative patients is that the *ANXA7* gene assay might provide a simple and reliable survival parameter for clinicians to include in patient management plans for early detection and treatment options. In prostate tumor specimens ANXA7 might provide a quick predictive/prognostic indicator to identify patients at risk of recurrent disease, warranting more-aggressive follow-up protocols or alternative treatment regimens [13].

8   ANXA7 as a Therapeutic Target in Cancer

Understanding the mechanism of ANXA7 function as a tumor suppressor can potentially allow for the development of therapeutic strategies for cancer treatment. Two general considerations make ANXA7 an attractive therapeutic target in cancer: it is selectively expressed in tumor cells and it is required for their viability. In dissecting the requisites for ANXA7 function, it was noticed that ANXA7 regulates IP3 Receptor expression and function. The importance of IP3 receptors in cancer cells is that IP3 Receptor activation by IP3 is the physiological stimulus needed to release calcium from the ER, thus triggering the mitochondrial permeability transition, which leads to apoptosis [66]. It is therefore possible that the action of the transfected *ANXA7* gene in tumor cells, may

be to elevate cytosolic Ca^{2+}, and to potentiate subsequent pro-apoptotic actions of the released calcium. The rationale of targeting ANXA7 for therapeutic intervention comes from experiments using a triple mutant ANXA7J in which putative calcium binding sites are mutated. ANXA7J is a dominant-negative antagonist of the calcium signaling pathway, and prevents the endogenous ANXA7 protein from binding to calcium. Loss of this calcium binding result in downregulation of IP3 receptors in cancer cells (unpublished data).

The anticancer properties of wild type ANXA7 have been recently tested in preclinical in vitro and in vivo experiments. Expression of wild type ANXA7 in vitro suppressed proliferation of prostate, breast and osteosarcoma cells by 95 % in a reaction associated with tumor cell apoptosis. To make it a potentially more-flexible therapeutic tool, we recently generated a replication-deficient adenovirus encoding wild type ANXA7. Administration of pAd-ANXA7 to nude mice prevented prostate and breast tumor formation, and suppressed the growth of existing tumors by 95 %. Gene therapy has recently emerged as a viable treatment in solid tumors by reestablishing apoptosis and checkpoint function with wild-type p53. A similar indication may be envisioned for the wild type ANXA7, either alone or in combination with chemotherapy. Finally, ANXA7 could provide an excellent target for cancer therapy because it also plays a crucial role in metastasis.

9   Future Directions

Since the identification of ANXA7 as a TSG, there has been considerable interest in ANXA7 from various viewpoints of biomedical research. The complexity of subcellular localization, the evidence for a dual role in both apoptosis and cell division, and the far-reaching consequences of dysregulated expression in cancer define the existence of a "ANXA7 pathway." The constituents of the pathway are only just now becoming known. Clearly, several questions on the biology of ANXA7 remain to be answered. The upcoming analysis of in vitro and in vivo transgenic models and the biochemical mapping of ANXA7 associated and regulated molecules using genomic and proteomic analysis will undoubtedly help the positioning of the ANXA7 pathway in the apoptosis cascade and the regulation of cell division. From a more translational perspective, the next set of challenges will be to integrate the ANXA7 pathway in the global genetic fingerprinting of cancer patients, linking individual patterns of gene expression to prediction of clinical outcome and response to therapy. Recent preclinical studies seem to "prove the principle" that ANXA7 is a viable therapeutic target in cancer, either directly or by affecting cell viability, and the first clinical trial using gene therapy for prostate cancer and

antisense ANXA7 for breast cancer may not be too far behind. As knowledge of the ANXA7 pathway grows, the knowledge of additional critical requisites for apoptosis inhibition or cell division will probably increase and, with that, more options for molecular diagnosis and therapeutic intervention will become available.

References

1. Shen WP, Young RF, Walter BN et al (1990) Molecular analysis of a myxoid chondrosarcoma with rearrangements of chromosomes 10 and 22. Cancer Genet Cytogenet 45: 207–215

2. Jenkins RB, Hay ID, Herath JF et al (1990) Frequent occurrence of cytogenetic abnormalities in sporadic nonmedullary thyroid carcinoma. Cancer 66:1213–1220

3. Morita R, Saito S, Ishikawa J et al (1991) Common regions of deletion on chromosomes 5q, 6q and 10q in renal cell carcinoma. Cancer Res 51:5817–5820

4. Shah NK, Wagner J, Santos G, Griffin CA (1992) Karyotype at relapse following allogeneic bone marrow transplantation for chronic myelogenous leukemia. Cancer Genet Cytogenet 61:183–192

5. Oberstrass J, Ring GU, Vogeley KT et al (1994) Allelieuntersuchungen auf Chomosom 10q21-26 in malignen Gliomen. Verh Dtsch Ges Pathol 78:413–417

6. Steck PA, Lignon AH, Cheong P et al (1995) Two tumor suppressive loci on chromosome 10 involved in human glioblastoma. Genes Chromosomes Cancer 12:255–261

7. Solic N, Collins JE, Richter A et al (1995) Two newly established cell lines derived from the same colonic adenocarcinoma exhibit differences in EGF-receptor ligand and adhesion molecule expression. Int J Cancer 62:48–57

8. Petersen S, Wolf G, Bockmuhl U et al (1998) Allelic loss on chromosome 10q in human lung cancer: association with tumour progression and metastatic phenotype. Br J Cancer 77:270–276

9. Piao Z, Park C, Park JH, Kim H (1998) Allotype analysis of hepatocellular carcinoma. Int J Cancer 75:29–33

10. Lacombe L, Orlow I, Reuter VE et al (1996) Microsatellite instability and deletion analysis of chromosome 10 in human prostate cancer. Int J Cancer 69:110–113

11. Li J, Yen C, Liaw D et al (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275:1943–1947

12. Maier D, Zhang Z, Taylor E et al (1998) Somatic deletion mapping on chromosome 10 and sequence analysis of PTEN/MMAC1 point to the 10q25-26 region as the primary target in low grade and high grade gliomas. Oncogene 16:3331–3335

13. Srivastava M, Bubendorf L, Nolan L et al (2001) ANXA7, a candidate tumor-suppressor gene for prostate cancer. Proc Natl Acad Sci U S A 98:4575–4578

14. Leighton X, Srikantan V, Pollard HB et al (2004) Significant allelic loss of *ANXA7* region (10q21) in hormone receptor negative breast carcinomas. Can Lett 210:239–244

15. Klee CB (1988) Ca2+-dependent phospholipid- (and membrane-)binding proteins. Biochemistry 27:6645–6653

16. Raynal P, Pollard HB (1994) Annexins: the problem of assessing the biological role for a gene family of multifunctional calcium-and phospholipid-binding proteins. BBA Biomembranes 1197:63–93

17. Smith PD, Moss SE (1994) Structural evolution of the annexin supergene family. Trends Genet 10:241–246

18. Flower RJ, Rothwell NJ (1994) Lipocortin-1: cellular mechanisms and clinical relevance. Trends Pharmacol Sci 15:71–76

19. McKanna JA (1995) Lipocortin 1 in apoptosis: mammary regression. Anat Rec 242:1–10

20. Rothhut B (1997) Participation of annexins in protein phosphorylation. Cell Mol Life Sci 53:522–526

21. Gerke V, Moss SE (2002) Annexins: from structure to function. Physiol Rev 82:331–371

22. Pollard HB, Rojas E (1988) Calcium acivated ANXA7 forms highly selective, voltage-gated channels in phosphatidylserine bilayer membranes. Proc Natl Acad Sci U S A 85: 2974–2978

23. Kubista H, Hawkins TE, Patel DR et al (1999) Annexin 5 mediates a peroxide-induced Ca(2+) influx in B cells. Curr Biol 9:1403–1406

24. Paweletz CP, Ornstein DK, Roth MJ et al (2000) Loss of annexin 1 correlates with early onset of tumorigenesis in esophageal and prostate carcinoma. Cancer Res 60:6293–6297

25. Chetcuti A, Margan SH, Russell P et al (2001) Loss of annexin II heavy and light chains in prostate cancer and its precursors. Cancer Res 61:6331–6334

26. Kang JS, Calvo BF, Maygarden SJ et al (2002) Dysregulation of annexin I protein expression in high-grade prostatic intraepithelial neoplasia and prostate cancer. Clin Cancer Res 8:117–123

27. Xin W, Rhodes DR, Ingold C et al (2003) Dysregulation of the annexin family protein family is associated with prostate cancer progression. Am J Pathol 162:255–261

28. Smitherman AB, Mohler JL, Maygarden SJ, Ornstein DK (2004) Expression of annexin I, II and VII proteins in androgen stimulated and recurrent prostate cancer. J Urol 171:916–920

29. Yadav AK, Renfrow JJ, Scholtens DM (2009) Monosomy of chromosome 10 associated with dysregulation of epidermal growth factor signaling in glioblastomas. JAMA 302:276–289

30. Bredel M, Scholtens DM, Harsh GR (2009) A network model of a cooperative genetic landscape in brain tumors. JAMA 302:261–275

31. Kataoka TR, Ito A, Asada H (2000) Annexin VII as a novel marker for invasive phenotype of malignant melanoma. Jpn J Cancer Res 91:75–83

32. Shirvan A, Srivastava M, Wang MG et al (1994) Divergent Structure of the human ANXA7 (annexin VII) gene and assignment to chromosome 10. Biochemistry 33:6888–6901

33. Zhang-Keck Z-Y, Burns AL, Pollard HB (1993) Mouse ANXA7 (annexin VII) polymorphisms and phylogenetic comparison with other ANXA7s. Biochem J 289:735–741

34. Zhang-Keck Z-Y, Srivastava M, Kozak CA et al (1994) Genomic organization and chromosomal localization of the mouse ANXA7 (annexin VII) gene. Biochem J 301:835–845

35. Srivastava M, Zhang-Keck ZY, Caohuy H et al (1996) Novel isoforms of ANXA7 in *Xenopus laevis*: multiple tandem PGQM repeats distinguish mRNA's in specific adult tissues and embryonic stages. Biochem J 316:729–736

36. Doring V, Schleicher M, Noegel AA (1991) Dictyostelium annexin VII (ANXA7). J Biol Chem 266:17509–17515

37. Gerke V (1991) Identification of a homologue for annexin VII (synexin) in Dictyostelium discoideum. J Biol Chem 266:1697–1700

38. Sun M-Z, Liu S, Tang J (2009) Targeting annexin A7 in hepatocarcinoma lymphatic metastasis. Chin J Lung Cancer 12:633–634

39. Gong X, Tang J, Geng X (2009) Expression and significance of Annex in A7 in gastric cancer and lymphatic metastasis. Int J Pathol Clin Med 29:369–373

40. Yang M, Liang Q (2011) Study the relationship between the expression of Annexin A7 and CT of nasopharyngeal carcinoma. J Chin Clin Med Imaging 22:6–9

41. Jimenez CR, Knol JC, Meijer GA (2010) Proteomics of colorectal cancer: overview of discovery studies and identification of commonly identified cancer-associated proteins and candidate CRC serum markers. J Proteomics 73:1873–1895

42. Srivastava M, Torosyan Y, Raffeld M (2007) ANXA7 expression represents hormonerelevant tumor suppression in different cancers. Int J Cancer 121:2628–2636

43. Guo C, Liu S, Greenway F, Sun M-Z (2013) Potential role of annexin A7 in cancers. Clin Chim Acta 423:83–89

44. Srivastava M, Bubendorf L, Raffeld M et al (2004) Prognostic impact of ANXA7-GTPase in metastatic and HER2 negative breast cancer patients. Clin Cancer Res 7:2344–2350

45. Srivastava M, Bubendorf L, Nolan L et al (2001) ANXA7 as a biomarker in prostate and breast cancer progression. Dis Markers 17:115–120

46. Bubendorf L, Sauter G, Moch H et al (1996) Prognostic significance of Bcl-2 in clinically localized prostate cancer. Am J Pathol 148:1557–1565

47. Silvestrini R, Veneroni S, Daidone MG et al (1994) The Bcl-2 protein: a prognostic indicator strongly related to p53 protein in lymph node-negative breast cancer patients. J Natl Cancer Inst 86:499–504

48. Villar E, Redondo M, Rodrigo I et al (2001) Bcl-2 expression and apoptosis in primary and metastatic breast carcinomas. Tumour Biol 22:137–145

49. Bonkhoff H, Remberger K (1996) Differentiation pathways and histogenetic aspects of normal and abnormal prostatic growth: a stem cell model. Prostate 28:98–106

50. Li P, Barraclough R, Fernig DG et al (1998) Stem cells in breast epithelia. Int J Exp Pathol 79:193–206

51. Feuerhake F, Sigg W, Höfter EA et al (2000) Immunohistochemical analysis of Bcl-2 and Bax expression in relation to cell turnover and epithelial differentiation markers in the nonlactating human mammary gland epithelium. Cell Tissue Res 299:47–58

52. Srivastava M, Montagna C, Leighton X et al (2003) Haploinsufficiency of ANXA7 tumor suppressor gene and consequent genomic instability promotes tumorigenesis in the ANXA7(+/−) mouse. Proc Natl Acad Sci U S A 100:14287–14292

53. Srivastava M, Atwater I, Glasman M et al (1999) Defects in IP$_3$ receptor expression, Ca^{2+}-signaling and insulin secretion in the ANXA7 (+/−) knockout mouse. Proc Natl Acad Sci U S A 96:13783–13788

54. Srivastava M, Kumar P, Leighton X et al (2002) Influence of the Anx7 +/− knockout mutation and fasting stress on the genomics of the mouse adrenal gland. Ann N Y Acad Sci 971:53–60

55. Kim T, Tao-Cheng JH, Eiden LE, Loh YP (2002) Large dense-core secretory granule biogenesis is under the control of chromogranin A in neuroendocrine cells. Ann N Y Acad Sci 971:323–331

56. Goping G, Pollard HB, Srivastava M, Leapman R (2003) Mapping protein expression in mouse pancreatic islets by immunolabeling and electron energy loss spectrum-imaging. Microsc Res Tech 61:448–456

57. Srivastava M, Eidelman O, Leighton X et al (2002) ANXA7 is required for nutritional control of gene expression in mouse pancreatic islets of Langerhans. Mol Med 8:781–797

58. Furuya Y, Lundmo P, Short AD et al (1994) The role of calcium, pH, and cell proliferation in the programmed (apoptotic) death of androgen independent prostatic cancer cells induced by thapsigargin. Cancer Res 54:6167–6175

59. Kyprianou N, Bains AK, Jacobs SC (1994) Induction of apoptosis in androgen-independent human prostate cancer cells undergoing thymineless death. Prostate 25:66–75

60. Wertz IE, Dixit VM (2000) Characterization of calcium release-activated apoptosis of LnCaP prostate cancer cells. J Biol Chem 275:11470–11477

61. Kass GE, Orrenius S (1999) Calcium signaling and cytotoxicity. Environ Health Perspect 107:25–35

62. Scoltock AB, Bortner CD, St J, Bird G et al (2000) A selective requirement for elevated calcium in DNA degradation, but not early events in anti-Fas-induced apoptosis. J Biol Chem 275:30586–30596

63. Lin XS, Denmeade SR, Cisek L, Isaacsm JT (1997) Mechanism and role of growth arrest in programmed (apoptotic) death of prostatic cancer cells induced by thapsigargin. Prostate 33:201–207

64. Tombal B, Weeraratna AT, Denmeade SR, Isaacs JT (2000) Thapsigargin induces a calmodulin/calcineurin-dependent apoptotic cascade responsible for the death of prostatic cancer cells. Prostate 43:303–317

65. Furuya Y, Ohta S, Ito H (1997) Apoptosis of androgen-independent mammary and prostate cell lines induced by topoisomerase inhibitors: common pathway of gene regulation. Anticancer Res 17:2089–2093

66. Szalai G, Krishnamurthy R, Hajnóczky G (1999) Apoptosis driven by IP (3)-linked mitochondrial calcium signals. EMBO J 18:6349–6361

Chapter 4

Experimental and Study Design Considerations for Uncovering Oncometabolites

Majda Haznadar and Ewy A. Mathé

Abstract

Metabolomics as a field has gained attention due to its potential for biomarker discovery, namely because it directly reflects disease phenotype and is the downstream effect of posttranslational modifications. The field provides a "top-down," integrated view of biochemistry in complex organisms, as opposed to the traditional "bottom-up" approach that aims to analyze networks of interactions between genes, proteins and metabolites. It also allows for the detection of thousands of endogenous metabolites in various clinical biospecimens in a high-throughput manner, including tissue and biofluids such as blood and urine. Of note, because biological fluid samples can be collected relatively easily, the time-dependent fluctuations of metabolites can be readily studied in detail.

In this chapter, we aim to provide an overview of (1) analytical methods that are currently employed in the field, and (2) study design concepts that should be considered prior to conducting high-throughput metabolomics studies. While widely applicable, the concepts presented here are namely applicable to high-throughput untargeted studies that aim to search for metabolite biomarkers that are associated with a particular human disease.

Key words Metabolomics, Biomarker discovery, Study design, Mass spectrometry, Oncometabolites, Analytical techniques

1 Introduction

Measurements from metabolomics experiments reflect genetic factors and are functional endpoints whose flux is measured in seconds (compared to minutes or even hours for proteins). In addition, metabolomics measurements also reflect the effect of environmental and lifestyle factors (e.g., diet, smoking), which are difficult to assess at the genome and proteome level. Such factors are assessed without making assumptions about what occurs at the genome and proteome level. Metabolomics is thus a promising tool that can detect perturbations associated with disease, even prior to clinical detection [1–3].

The characterization of D-2-hydroxyglutarate (D-2HG) production marks one of the earliest examples of oncometabolites. D-2HG levels are increased upon gain-of-function mutations of isocitrate dehydrogenase (IDH) in leukemias, gliomas, and other cancer types. High D-2HG levels lead to a competitive inhibition of α-ketoglutarate-dependent (α-KG-dependent) dioxygenase enzymes and posttranslational modification of proteins [4, 5]. Similarly, loss-of-function mutations in fumarate hydratase (FH) and succinate dehydrogenase (SDH), lead to an inhibition of fumarate and succinate production, respectively [6]. These were the first pieces of evidence directly linking altered metabolism and cancer, and such discoveries have led to a momentous and renewed interest in the field of metabolomics as a platform for discovery of robust metabolite markers of disease phenotypes and outcomes.

2   Specific Biological Questions Influence Utilized Approaches

2.1   NMR vs. Mass Spectrometry-Based Metabolomics (UPLC-MS, GC-MS)

The most utilized analytical techniques used in metabolomics are nuclear magnetic resonance (NMR) and mass spectroscopy (MS), which is often coupled to either liquid or gas chromatography (LC-MS or GC-MS, respectively) to separate the biological fluid before detection. Other techniques exist as well, such as capillary electrophoresis mass spectrometry. Use of specific techniques depends on the purpose of a given experiment. NMR, for example, does not require any pretreatment of a sample prior to the analysis, and as such, does not lead to damage of the analytes. Both NMR and MS techniques provide a wide range of metabolites in a single measurement. However, NMR is more reliable for providing exact concentrations of measured molecules. MS, on the other hand, is more sensitive and can detect metabolites that are below the limit of detection of NMR. Both techniques can be used to determine metabolite structures, but NMR's sensitivity is far too low for low-abundance metabolites and lacks a separation component, making LC-MS and GC-MS superior for discovery of biomarkers and potential oncometabolites. Currently, NMR is reserved for studying known and high-abundance targets, or for identifying unknown metabolites generated by MS approaches. Therefore, the two techniques can be complimentary.

Metabolites are chemically diverse, not encoded in the genome, and comprise carbohydrates, amino acids, lipids, and nucleotides among others. Because of this diversity, extraction, separation, and analytical techniques used depend on the specific experiment. Furthermore, untargeted approaches leverage unbiased techniques that will generate the most widespread types of captured metabolites, whereas targeted approaches are optimized to best quantitate a metabolite or metabolite type in question. Additional details regarding procedures for measurement of specific classes of

molecules are described in the next sections. For the time being, however, LC-MS seems to be one of the most preferred metabolomic analytical platforms in metabolomics. Because of the widespread use of MS in biomarker research, we focus on this technology throughout this chapter.

2.2 Types of Chromatography

Depending on the biospecimen being analyzed, hydrophobicity of the majority of the metabolites it comprises should be considered. For example, aqueous urine will contain a large number of hydrophilic compounds, while blood (plasma, serum) will contain a large number of hydrophobic compounds, such as lipids. Based on this information, a researcher can decide what type of chromatography should be coupled with MS to attain a better retention of a targeted class of molecules.

Hydrophilic Interaction Chromatography (HILIC) is similar to normal phase liquid chromatography but with an aqueous mobile phase. This chromatography leads to an excellent retention and separation of polar metabolites. This type of chromatography is typically chosen to study aqueous biospecimens, such as urine. Some typical analytes of HILIC are amino acids, nucleobases (purines and pyrimidines), nucleosides, and carbohydrates. In HILIC, most of the hydrophobic molecules will elute early in the run, and will therefore not be properly retained. Thus, for the study of hydrophobic compounds, a different type of chromatography should be utilized.

Reverse Phase (RP) chromatography is the opposite of normal phase, resulting in the adsorption of hydrophobic molecules onto a hydrophobic solid support in a polar mobile phase. Therefore, the more hydrophobic the molecule, the more time it will spend on the solid support. Opposite of HILIC, most of the hydrophilic metabolites will elute early and will not be sufficiently retained. RP is often chosen to study metabolites in biospecimens such as serum and plasma, which are enriched for compounds such as lipids. However, RP chromatography is the most popular separation technique for untargeted approaches, because it applies to a wide range of molecules including charged and polar molecules. It also allows precise control of variables such as organic solvent type and concentration, pH and temperature. Furthermore, RP columns are usually efficient and stable, making them suitable for analyzing a wide class of molecules with different retention times.

2.3 Lipidomics

Although technically a study of hydrophobic compounds, lipidomics has emerged as a distinct subdiscipline. A large portion of the metabolomics research community focuses specifically on studying lipids. Lipids are extraordinarily diverse in their chemical structures, polarities, and potential to ionize. LC-MS lipidomics is remarkably robust, high-throughput, qualitative, and semiquantitative, and is successfully applied to study many different diseases.

This approach relies on high-resolution/accurate-mass (HR/AM) and high-energy collisional dissociation (HDC) fragmentation of an orbitrap LC-MS. Specialized deconvolution software that can distinguish specific lipid molecules by their unique fragmentation patterns is then applied to identify individual metabolites. Many of the lipids are already catalogued and therefore present an opportunity for a faster identification of unknown metabolites in global lipidomics studies.

3   Biospecimens and Sample Processing Procedures

One of the significant benefits of metabolomic profiling is minimal sample processing, ensuring minimal disturbance of the metabolites being measured. In order to keep consistency and increase reproducibility, it is important to keep in mind that standard operating procedures (SOPs) should be in place and strictly followed. Information such as the time and duration of collection should be recorded so that these potential confounders can be adjusted during data analysis. Samples should be frozen as soon as possible upon collection at –80 °C and shipped on dry ice. Prior to processing, samples should be placed at 4 °C to thaw overnight.

Urine and plasma require minimal volume, as they are usually significantly diluted (1:4 to 1:20 for global analysis and depending on the sensitivity of the instrumentation) with organic materials such as aqueous methanol or acetonitrile. Samples are then vortex-mixed with an organic "crash solution" of choice (e.g., acetonitrile, methanol), and centrifuged at maximum speed for about 15 min at 4 °C in order to precipitate out the proteins. This centrifugation step is also important to avoid clogging of the columns by residual proteins. For tissue samples, tissue pieces as small as 5 mg can be cryogenically grinded at very low temperatures, allowing for an isolation of measurable metabolites, while protecting those that may be temperature-sensitive. For example, the Cryomill instrument (Retsch, Inc.) can be used, which pulverizes tissue samples with stainless steel balls at high frequencies of vibration and temperature as cold as –200 °C by constant influx of liquid nitrogen. If extracting metabolites from cell cultures, cell metabolism should be quenched first, in order to stop it immediately. Methanol is commonly used for quenching. Usually, a minimum of two million cells is required to capture the cell metabolome at an MS-detectable level. Cells should be sufficiently lysed for metabolite extraction, and this can be accomplished either by several freeze–thaw cycles, using water before organic crash buffers, or by utilizing sonication.

Specific organic crash solutions should be considered depending on the goal of the study. In order to profile the widest class of molecules (both polar and nonpolar), aqueous methanol (e.g., 80% methanol, 20% water) has been proven to be a successful and appropriate crash solution. If attempting to enrich for lipids, as

well as capturing polar compounds, a mixture of ice-cold chloroform–methanol–water (e.g., 2:5:2, v:v:v) can be used for metabolite extraction.

Importantly, it is necessary to consider the effect of multiple freeze/thaw cycles on the metabolite measurements in biofluids such as urine and blood. In LC-MS, GC-MS, and NMR studies, samples clustered similarly together even after several freeze–thaw cycles [7]. However, levels of carnitine in blood were shown to decrease as the number of freeze–thaw cycles increases [8]. While these results suggest that samples having undergone several freeze–thaw cycles are still appropriate to use in large-scale metabolomics studies, it is best to be cautious and use samples with minimal freeze–thaw cycles. Furthermore, it is recommended to use samples that have the same number of freeze–thaw cycles to avoid possible confounding. Finally, when collecting samples for metabolomics analysis, it is ideal to generate multiple aliquots of the same specimen, thereby allowing the same freeze–thaw cycle specimen to be utilized for different analysis (e.g., pilot experiment, technical validation experiment).

4   Quantitation of Metabolites of Interest

4.1   Methodology   The field of metabolomics is still concerned with the number of metabolites present in a given system or biofluids, as well as their characterization. Upon overcoming this hurdle, metabolomics should move into a stage of high-throughput quantitation of thousands of metabolites using tandem mass spectrometry. Liquid-chromatography tandem mass spectrometry or triple quadrupole mass spectrometry can be leveraged for untargeted and targeted metabolomics, respectively. It is evident that the future of metabolomics is intimately interwoven with the development of sensitive instrumentation and software that is capable of quickly and accurately processing thousands of molecules in exploratory studies, so that routine analysis can be conducted on important targets related to a phenotype under investigation.

In the case of LC-MS, thousands of mass to charge ratios (m/z) with corresponding retention times (RTs) are captured through global metabolomics. A critical first step in the data processing is data de-convolution, where the goal is to create unique spectra for each metabolite, thereby separating co-eluting metabolites from each other. Software such as Marker Lynx (Waters, Inc.), XCMS [9], MAVEN [10], PRIME, and others have been designed to perform deconvolution. Furthermore, de-isotoping techniques are applied (by the same software) to remove fragments that pertain to the same whole molecule from the analysis. Next, appropriate statistical methods are applied to uncover metabolites of interest to the study (e.g., most robust predictors of a disease status). Finally, in order to assign and verify identities of the metabolites of

interest, fragmentations of the markers are carried out using triple quadrupole mass spectrometry. Currently, databases are being constructed that contain accurate masses with putative identities, and in some cases chromatographic spectra. These databases help inform the researcher which high-purity standards can be purchased commercially or which should be synthesized (if no standard exist). Standards are ultimately required to assign identities to unknown metabolites.

4.2   Databases   While many compounds can be detected in a given experiment, they are not as easily identified. Researchers have made great progress in that regard, with the development of a freely accessible web-based data repository, METLIN, that was developed to facilitate the identification of a broad array of unknown metabolites through mass analysis [11]. Other databases that can facilitate identification of the unknown metabolites include Human Metabolome Database [12], BinBase [13], and others. However, the most significant difficulty arises when it is impossible to run a commercially available standard against the unknown metabolite of interest to compare the spectra and fragmentation patterns for the identification. Without a commercially available compound, a standard has to be synthesized. Such problems can be discouraging to the research community in terms of conducting global metabolomics, and researchers often opt for targeted approaches with specific biological questions in mind. The importance of having extensive equipment and well-trained researchers to take on such large tasks is paramount.

5   Study Design Considerations

Prior to starting a large metabolomics experiment, it is essential to consider what platform should be applied (discussed above) as well as what and how many samples should be included to extract meaningful results (e.g., potential biomarkers). Importantly, the large dynamic range of metabolite abundances (up to 10^{14}), sample to sample variation, and instrumentation variation make the challenges of designing robust metabolomics studies unique. To ensure quality data and a well-designed study (and therefore meaningful results), it is always best to involve a metabolomics data specialist or statistician prior to collecting any samples or data.

6   Assessing Variation and Sample Size

With thousands of peaks being generated for one experiment, each having potentially very large dynamic ranges, it is important to get a sense of metabolite variability and consequently the sample size

that should be employed to attain meaningful results. Conducting a pilot study is arguably the best way of assessing variability, and it will also allow for optimization of experimental conditions. Assessing publicly available data, from databases such as Metabolights [14] or Metabolomics Workbench ("The Metabolomics Workbench, http://www.metabolomicswork-bench.org/") could also help with estimating variability, although it is essential to keep in mind that sample type and acquisition can affect variability.

Pilot experiments are typically performed on a small number of samples (e.g., 10–50 controls and 10–50 diseased samples). Standard deviations and means for each metabolite can be calculated and from these, the sample size required to attain a specified statistically significant p-value (e.g., 0.01) and fold change (e.g., 1.5-fold) can be calculated. Software such as the University of Vanderbilt's PS Software (http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize) can be utilized to calculate required sample size for desired outcome.

7   Including Proper Technical Controls

To ensure quality of the overall experiment and the reliability of the measurements, it is important to consider including technical controls when designing an experiment. First, pooled samples should be run randomly throughout the experiment. For example, an aliquot from several samples could be combined in one large pooled sample and processed randomly throughout the run. The extracted abundances of these samples should be very similar. Second, a portion of the samples (e.g., 10%) should be randomly selected and processed in duplicate. Similarly to the pooled samples, metabolite abundances for these should be strongly correlated. Third, high-purity stable isotope-labeled internal standards can be injected in the samples and aid in the identification of metabolites of interest. The use of isotope-labeled internal standards, while costly, is overall an excellent choice for improved quantification and identification [15, 16]. Fourth, blanks could be run to assess sample carryover. However, blanks can cause shifts in chromatograms because of the disruption in the sample types (e.g., blanks vs biofluid samples). These shifts can be mitigated by equilibrating the column with multiple pooled samples from the same biofluid run in the experiment, after blanks are run.

Finally, the order in which the samples are injected should be randomized throughout the experiment. This step is critical for avoiding possible confounding of the sample injection order. Also, should any issues arise later in the run (e.g., sample carryover observed starting at sample 100), samples run up to the observed issue could still be salvaged.

8 Assessing Experiment Quality

First and foremost, it is essential to evaluate raw chromatograms one by one and ensure that there are no obvious red flags. Errors to look out for include overall low signal, interruption of signal, or unusual spikes (which could indicate sample contamination or potentially unwanted xenobiotics). Next, once the raw signal abundances are extracted, normalized, and scaled (see below for discussion on normalization and scaling), the following basic statistical and visualization techniques can be applied to assess the reliability of the measurements. First, correlation coefficients can be calculated to evaluate the correlation of metabolite abundances within pooled and duplicate samples. These coefficients should be very high (>0.9). Second, to compare the variation in metabolite abundances between disease, non-disease, and technical control (e.g., pooled samples and duplicates) samples, coefficients of variation (CVs) can be calculated. CVs are simply standard deviations normalized by the mean, thereby enabling standard deviations to be compared between different groups. The CVs should be demonstrably larger in the biological samples than in the technical control samples. Finally, clustering techniques can be applied to obtain a global picture of how similar the samples are to each other. Hierarchical clustering can be applied for this purpose. Such clustering results in a dendrogram, which groups samples together according to a user-defined distance metric (e.g., correlation). High-dimensionality reduction techniques such as principal components analysis (PCA) and multidimensional scaling are also commonly applied. These techniques result in a two- or three-dimensional plot, where the distance between points (e.g., samples) reflects the similarity between samples.

Analagous to sample quality assessment, the reproducibility and quality of metabolite abundances can be evaluated. For example, CVs can be calculated for each metabolite across identical quality control samples (e.g., pooled samples) and metabolites with high CVs (e.g., 30 %) can be removed prior to normalizing and further processing. In addition, metabolites with abundance levels close to noise can be removed from further analysis. For instance, metabolites with abundances lower than the 10th percentile of all mean metabolite abundance could be regarded as noise.

9 Normalization

The overall goal of normalization is to ensure that samples are comparable to each other by reducing fluctuations in metabolite abundances that are not due to the biological outcome in question (e.g., differences in abundances between disease groups). Technical variation in measurements can be due to differences in sample preparation, ionization efficiency, and instrument conditions, for

example. Several techniques can be applied to reduce this unwanted variation. One technique is "total ion current" (TIC) normalization [17], where the sum of all signal abundances in a sample is made equal across all the samples. For each sample, signal abundances are divided by the total sum of abundances, then multiplied by a constant (e.g., mean sum of abundances of all samples). Another method, "MS total useful signal" (MSTUS) normalization, was introduced more recently and builds upon TIC normalization [18]. Rather than using all the signal abundances as in TIC normalization, only signals that are common to all or to the large majority of samples are retained. MSTUS normalization thus removes potential artifacts (e.g., drug intake) that may artificially skew the sum of abundances. Another option is to normalize signal abundances to that of spiked-in standards (with known amounts) but care should be taken to include standards that elute at different times. In urine, "housekeeping" metabolites (e.g., creatinine) have been applied yet may not be recommended since abundances of creatinine values are actually quite variable. Finally, abundances can be normalized to sample tissue weight/volume, or osmolality, or other physical measurements of the starting material.

In addition to normalization, metabolite abundances are typically transformed or scaled. This step is important for downstream statistical analysis, which oftentimes requires the data to fit a specific distribution (e.g., normal). Typical transformations include applying log or power-transformations, and typical scaling include mean-centering or Pareto scaling.

10   Validation

To ensure that the results are reproducible, it is essential to validate findings. Two types of validation are important to consider: technical and biological. Technical validation evaluates whether the abundances obtained are reproducible using the same or other instrumentation. For example, a subset or all the samples could be run on the same instrument in a different laboratory or on a different instrument (e.g., LC-MS and QTOF-MS). In both cases, observing similar associations between metabolites and disease is the desired outcome and would rule out potential technical artifact. Furthermore, calculating intraclass correlation coefficients (ICCs) is useful and gives a measurement of reproducibility. For biomarkers to be considered as clinically useful, ICCs greater than 0.7 are required. Biological validation of the results is equally important and requires running a second experiment on a new set of samples that are preferably not part of the same cohort. This type of validation will ensure that the findings are not specific to the population under study and is very useful in evaluating the global utility of the uncovered biomarkers.

11 Drawbacks

While the need to use complex data-interpretation techniques and combinations of analytical methods is not ideal, metabolomics may allow for the closest examination of causal relationships between genetics, environmental exposures, and disease phenotypes and outcomes. Although it is difficult to predict the number of metabolites produced by any given system, in comparison to the number of genes for example, this problem is not insurmountable and is similar to other fields, such as epigenetics. Other challenges in the field include a lack of well annotated central databases with spectral data from various biospecimens, and a standardization of the methods being conducted across laboratories. Currently, it is very difficult to directly compare results between different methods, laboratories, and different specimens. There is also a very wide concentration dynamic range, that can be as high as 10^{14}, and presence of many significantly low-abundance metabolites. The sources of variability are a frequent problem: sample, analytical methods, workflow, reagents, etc. In order to minimize the variability and discover valuable signals to illuminate questions at hand, it is of the outmost importance to conduct proper study designs with appropriate controls, randomized sampling, consistent sample preparation and storage.

Many of the aforementioned issues are common to other high-throughput biochemical analyses. Nevertheless, global profiling studies can be very rewarding, as the set of metabolites present in a cell or an organism at a given time point offers a unique, downstream snapshot of the physiological state of the system.

References

1. Dunn WB, Broadhurst DI, Atherton HJ et al (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev 40:387–426

2. van der Greef J, Stroobant P, van der Heijden R (2004) The role of analytical sciences in medical systems biology. Curr Opin Chem Biol 8:559–565

3. Stepien M, Duarte-Salles T, Fedirko V et al (2016) Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: findings from a prospective cohort study. Int J Can 138:348–360

4. Dang L, White DW, Gross S et al (2010) Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 465:966

5. Ward PS, Patel J, Wise DR et al (2010) The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. Cancer Cell 17:225–234

6. Pollard PJ, Briere JJ, Alam NA et al (2005) Accumulation of Krebs cycle intermediates and over-expression of HIF1alpha in tumours which result from germline FH and SDH mutations. Hum Mol Genet 14:2231–2239

7. Robaglia A, Cau P, Bottini J, Seite R (1989) Effects of isolation and high helium pressure on the nucleolus of sympathetic neurons in the rat superior cervical ganglion. J Auton Neurosci 27:207–219

8. Yin P, Peter A, Franken H et al (2013) Preanalytical aspects and sample quality assess-

ment in metabolomics studies of human blood. Clin Chem 59:833–845

9. Libiseller G, Dvorzak M, Kleb U et al (2015) IPO: a tool for automated optimization of XCMS parameters. BMC Bioinformatics 16:118

10. Clasquin MF, Melamud E, Rabinowitz JD (2012) LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Curr Protoc Bioinform. Chapter 14:Unit14. 11. doi:10.1002/0471250953.bi1411s37.

11. Smith CA, O'Maille G, Want EJ et al (2005) METLIN: a metabolite mass spectral database. Ther Drug Monit 27:747–751

12. Wishart DS, Jewison T, Guo AC et al (2013) HMDB 3.0—the human metabolome database in 2013. Nucleic Acids Res 41:D801–D807

13. Skogerson K, Wohlgemuth G, Barupal DK, Fiehn O (2011) The volatile compound BinBase mass spectral database. BMC Bioinformatics 12:321

14. Haug K, Salek RM, Conesa P et al (2013) MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 41:D781–D786

15. Beynon RJ, Pratt JM (2005) Metabolic labeling of proteins for proteomics. Mol Cell Proteomics 4:857–872

16. Rousseaux M, Petit H, Hache JC et al (1985) Ocular and head movements in infarctions of the thalamic region. Rev Neurol 141:391–403

17. Fung ET, Enderwick C (2002) ProteinChip clinical proteomics: computational challenges and solutions. Biotechniques Suppl:34–38, 40–41

18. Warrack BM, Hnatyshyn S, Ott KH et al (2009) Normalization strategies for metabonomic analysis of urine samples. J Chromatogr B Analyt Technol Biomed Life Sci 877: 547–552

Chapter 5

Targeting Deubiquitinating Enzymes and Autophagy in Cancer

Ashley Mooneyham and Martina Bazzaro

Abstract

Maintenance of proper cellular homeostasis requires constant surveillance and precise regulation of intracellular protein content. Protein monitoring and degradation is performed by two distinct pathways in a cell: the autophage–lysosome pathway and the ubiquitin–proteasome pathway. Protein degradation pathways are frequently dysregulated in multiple cancer types and can be both tumor suppressive and tumor promoting. This knowledge has presented the ubiquitin proteasome system (UPS) and autophagy as attractive cancer therapeutic targets. Deubiquitinating enzymes of the UPS have garnered recent attention in the field of cancer therapeutics due to their frequent dysregulation in multiple cancer types. The content of this chapter discusses reasoning behind and advances toward targeting autophagy and the deubiquitinating enzymes of the UPS in cancer therapy, as well as the compelling evidence suggesting that simultaneous targeting of these protein degradation systems may deliver the most effective, synergistic strategy to kill cancer cells.

Key words Deubiquitinating enzymes, Autophagy, Cancer, Small molecule inhibitors, Ubiquitin–proteasome system

1 Introduction to Protein Degradation Systems in The Eukaryotic Cell

Maintaining proper cellular protein homeostasis requires constant surveillance of protein quality and quantity coupled with tightly regulated protein degradation systems [1]. Misfolded, damaged, or unnecessary proteins are continuously broken down to minimize dangerous intracellular proteotoxic stress. The amino acids resulting from the protein degradation process can then be recycled and incorporated into new proteins, which completes the homeostatic protein cycle. In eukaryotic cells, two major pathways regulate protein degradation: the ubiquitin–proteasome pathway and the autophage–lysosome pathway [2]. In general, the ubiquitin–proteasome system is used to degrade short-lived proteins and the autophage–lysosome system is used to degrade long-lived proteins. Protein degradation pathways are frequently dysregulated in multiple cancer types and can be both tumor

suppressive and tumor promoting [3, 4]. This knowledge has presented the ubiquitin–proteasome system (UPS) and autophagy as attractive cancer therapeutic targets [5, 6].

1.1 The Autophage–Lysosome System

Three types of autophagy have been identified: chaperone-mediated autophagy, microautophagy, and macroautophagy. Molecular chaperones in the cytosol and the lysosomal lumen control chaperone-mediated autophagy. Microautophagy is distinct due to the direct engulfment of target proteins by the lysosome for degradation [7]. In this chapter, macroautophagy will be focused on given that nearly all work related to autophagy and cancer development, progression, and therapy refers to macroautophagy [8].

Macroautophagy (referred to as autophagy for the remainder of this chapter) involves the creation of double-membraned vesicles called autophagosomes to enclose small portions of the cytoplasm. Autophagosomes can then fuse with a lysosome containing digestive enzymes. This fusion allows the contents of autophagosomes to be digested in a contained environment. Autophagy is critical for cellular survival under stressed conditions such as nutrient starvation, because digestion of less-essential self-components provides a source of metabolic building blocks used to refuel cells. This type of macroautophagy is commonly referred to as induced autophagy. Another common type of macroautophagy is basal autophagy. Under normal cellular conditions, basal autophagy is an essential degradation pathway for removing damaged organelles and protein aggregates from the cytosol [9].

1.2 The Ubiquitin–Proteasome System

The hallmark of the ubiquitin–proteasome system (UPS) is its use of the 76-amino acid protein ubiquitin to selectively mark proteins for targeted degradation. Proteins tagged with a polyubiquitin chain are recognized and degraded by the 26S proteasome of the UPS. Prior to degradation, the ubiquitin tags are removed from the target protein by deubiquitinating enzymes (DUBs) which cleave the tags into ubiquitin monomers that can then be recycled within the cell to tag subsequent proteins for degradation. Once the polyubiquitin chain is removed, the target protein substrate can be threaded into the core of the 26S proteasome, where it is unfolded and hydrolyzed. Peptides from the degraded protein exit the bottom of the complex, where they are then degraded into amino acids used to synthesize new proteins in a continuous process.

1.3 Cross Talk Between Autophagy and UPS

Until recently, UPS and autophagy protein degradation pathways were thought to act in parallel with each other to prevent the accumulation of polyubiquitinated and aggregated proteins. However, accumulating evidence suggests some cross talk between the ubiquitin–proteasome pathway and the autophage–lysosome pathway [10, 11]. For example, although autophagy is generally understood to be responsible for degradation of long-lived proteins, it

can compensate for degradation of short-lived ubiquitinated proteins when UPS activity is impaired. The UPS can also act in a compensatory manner, with proteasomal degradation shown to be upregulated in conditions when autophagy-mediated protein degradation is compromised. Both autophagy and UPS have recognized importance in maintaining proper health and homeostasis of eukaryotic cells. To inhibit either of these pathways greatly cripples a cell's ability to eliminate dangerous intracellular proteotoxic stress, making them a tantalizing target in cancer therapy due to the already high levels of proteotoxic stress existing inherently in cancer cells. This chapter will discuss results of targeting each system individually in cancer and the novel strategy of targeting both systems simultaneously to increase the efficacy of protein degradation inhibitors.

2   Autophagy and Cancer

It has become apparent that autophagy is a critical pathway in tumor development and cancer therapy [12, 13]. However, conflicting results have been published regarding autophagy's precise function in cancer. Is autophagy tumor suppressive or tumor promoting? The current consensus suggests both [14]. In normal cells, autophagy can act as a tumor suppressor by degrading damaged organelles and aggregated proteins before they can accumulate [15]. Preventing this accumulation halts tumor progression in its early stages by protecting normal cells from oxidative stress and DNA damage that could induce transformation into tumor cells. Despite this protective role, autophagy is also well known to be upregulated in cancer cells to promote tumor growth [16].

In cancer cells, autophagy upregulation is implemented as a survival mechanism. The rapid proliferation rate of cancer cells results in high protein turnover and increased metabolic demands [17]. These challenges are compounded by the fact that cancer cells are often deprived of nutrients due to their location within a tumor [18]. This leaves a threefold opportunity for autophagy to be pro-tumorigenic: autophagy can compensate for the high protein turnover by degrading excess proteins and compromised organelles, then the protein and organelle degradation products can be recycled to overcome nutrient-poor environments and supplement the cancer cells an alternate source of metabolic and synthetic substrates. Inhibition of autophagy is therefore an attractive target for cancer therapy. By inhibiting autophagy function in autophagy-dependent cancer cells, tumor cell death will be accelerated at a rate much greater than normal cells [19].

Chloroquine and hydroxychloroquine are FDA-approved autophagy inhibitors commonly used to treat malaria. New interest in autophagy as a cancer therapy has repurposed these malarial

autophagy inhibitors as anticancer therapeutics in recent clinical trials, especially within the context of combinational treatment with conventional chemotherapy, radiotherapy, and immunotherapy regimens [20]. Hydroxychloroquine/chloroquine enters lysosomes and halts protein degradation by reducing lysosomal acidity, which effectively blocks autophagy in the last step. Alternatively, there are inhibitors such as HDAC6 inhibitors that prevent the fusion of lysosomes and autophagosomes [21]. This also is an effective mechanism of blocking autophagic protein degradation by preventing the protein exposure to degradative enzymes.

In most preclinical and clinical trials, autophagy inhibitors are used in combination with current chemotherapy regimens rather than as a single agent [22]. Since chemotherapy has been shown to induce autophagy in cancer cells as a possible tumor-protective mechanism, inhibiting autophagy may resensitize cancer cells to the chemotherapy treatment. Most preclinical trials show promising results and support autophagy inhibitors as a cancer therapy option [23]. These results are especially profound in cancers such as pancreatic cancer that are "addicted" to autophagy for tumorigenic development in addition to relying on autophagy in stressed conditions [19].

3 Ubiquitin Proteasome System and Cancer

Cancer cells sustain survival by aberrant expression of proteins that are central to cell cycle and apoptosis. If the proteins are involved in suppressing apoptosis and promoting cell cycle progression, they are often overexpressed in cancer; if the proteins are involved in cell cycle checkpoints and induction of apoptosis, they are often underexpressed in cancer [24]. Due to their high requirement for metabolic activity and protein turnover, cancer cells are extremely dependent on proteasome degradation. This understanding led to the idea of using proteasome inhibition as a cancer treatment [25]. The goal of proteasome inhibition is to cause extensive and rapid accumulation of intracellular proteins that would prove toxic to cancer cells. As expected, preclinical studies and clinical evidence has shown proteasome inhibition does indeed exhibit anticancer effects both in vitro and in vivo [26, 27].

The most notable success in the field of proteasome inhibition as cancer therapy is the small molecule proteasomal inhibitor bortezomib. Just five years after submission of its Investigational New Drug Application, bortezomib was granted accelerated approval by the FDA in 2003[28]. It is primarily used to treat hematologic malignancies such as multiple myeloma and mantle cell lymphoma. Bortezomib reversibly inhibits the 26S proteasome and is initially incredibly effective at increasing survival of patients with hematological malignancies such as multiple myeloma [26]. However, these

successes were consistently met with onset of bortezomib resistance and the anticancer activity of bortezomib failed to translate to solid tumor setting. Consequently, bortezomib provided evidence and rationale for targeting the proteasome in cancer therapy, but it was clear that its considerable limitations warranted further experimentation and development of novel ways to inhibit the ubiquitin proteasome system.

Carfilzomib is a second generation proteasome inhibitor that sports one major difference to bortezomib: rather than reversibly inhibit, carfilzomib irreversibly inhibits the 26S proteasome. The intent behind the design was to increase potency of this drug and prevent resistance from occurring [29]. Similar to bortezomib, its effects in hematologic malignancies were robust and undeniable. Although tolerated in patients with solid malignancies, it has also struggled to amount significant antitumor activity in advanced solid tumors [30]. Notable third-generation inhibitors of the proteasome include marizomib, ixazomib, oprozomib, and delanzomib. Third-generation inhibitors are not yet FDA approved and are currently undergoing clinical trials. The above-mentioned third-generation proteasome inhibitors have all shown preclinical efficacy and dose tolerance in settings such as multiple myeloma [31].

Despite having first, second, and now third-generation proteasome inhibitors in clinical trials, all have suffered class-wide limitations. Proteasome inhibition only targets the last step of the UPS and often results in the development of dangerous side effects such as peripheral neuropathy and resistance in patients. To overcome the resistance to proteasomal inhibitors seen acquired in hematological malignancies and inherently in solid malignancies, as well as to improve upon common toxicities associated with proteasomal inhibition, interest has developed into targeting other essential components of the UPS that may provide an alternative strategy to reduce highly proliferative or apoptosis-resistant tumor cells to a more vulnerable state. Most recently, interest in development of inhibitors against deubiquitinating enzymes (DUBs) has taken off as a novel approach to targeting the UPS [32].

4   Deubiquitinating Enzymes and Cancer

A valid alternative for targeting the UPS in cancer involves targeting DUBs. DUBs are a family of proteases consisting of at least 98 individual DUBs identified to date [33]. In an already complex UPS system, DUBs add another level of regulation. DUBs are responsible for several functions including: liberation of ubiquitin from protein substrates, editing of polyubiquitin signal on protein substrates to change the fate of the protein, disassembling polyubiquitin chains to free ubiquitin monomers, and cleaving ubiquitin

precursors or adducts to regenerate active ubiquitin [34]. These functions can work to both encourage degradation or prevent degradation of proteins, and as such reinforce the significant role DUBs play in the UPS and in maintenance of protein homeostasis [35]. There are five DUB subfamilies identified: ubiquitin C-terminal hydrolases (UCHs), ubiquitin-specific proteases (USPs), ovarian tumor proteases (OTUs), Josephins, and JAB1/MPN/MOV34 metalloenzymes (JAMM/MPN+) [36].

The UCH and USP subfamilies are by far the most represented members of the DUB family, comprising greater than 90 % of the mammalian cell DUBs pool. Members of both USP and UCH families are differentially expressed and activated in a number of cancer types [37]. Notably, their aberrant activity has been linked to cancer progression, onset of chemoresistance, prognosis, and clinical outcome [38]. It has become recently clear that different cancers upregulate specific DUBs, which provides the added potential advantage of creating targeted therapy specific to each cancer through carefully selected DUB inhibition. Therefore, DUBs have been suggested as potential therapeutic targets for cancer treatment. Currently, UCH and USP are the only subfamilies for which small molecule inhibitors have been developed and characterized.

Dedicated efforts to design and synthesize small molecule inhibitors against cancer-related DUBs have intensified over the past five years. The majority of DUBs within a cell are cytosolic, with the exception of USP14 and UCH37/UCHL5 which are directly associated with the 26S proteasome [39]. Recent compounds synthesized to inhibit DUBs as cancer therapy include RA-190 and WP1130. The molecule RA-190 covalently binds the RPN13 ubiquitin receptor, which has been shown to be frequently amplified in ovarian and colon cancers, thereby disrupting recognition of ubiquitinated protein substrates for degradation by the UPS [40]. WP1130 acts as a partially selective DUB inhibitor targeted against USP9X, USP5, USP14, and UCH37/UCHL5 [41]. Unfortunately, although WP1130 was effective at inducing cell death in multiple myeloma and mantle cell lymphoma, its poor solubility and pharmacokinetic properties prevented further clinical development of this drug. After extensive screening and analysis, two promising derivatives of WP1130 have emerged: G9 and YJ-8-067. Both selectively target USP9X and have improved antitumor activity and superior in vivo performance to WP1130 in preclinical studies [42].

Notable proteasome-associated DUB inhibitors include b-AP15, VLX1570, and RA-9. Small molecule inhibitors VLX1570, and RA-9 have increased potency and solubility compared to b-AP15; however, all have shown anti-multiple myeloma activity. This anti-multiple myeloma activity extends to multiple

myeloma cells that are resistant to proteasome inhibition, which further supports DUB inhibition as a strategy to overcome proteasome inhibitor resistance [43, 44]. Strikingly, RA-9 has additionally been shown to exert antitumor activity without apparent side effects in preclinical solid tumor models such as breast and ovarian cancer [45]. The preliminary success of these compounds supports the rationale for further study centered on DUB activity inhibition to overcome resistance to UPS inhibiting drugs that target the proteasome. DUB inhibitors also hold promise to finally translate the antitumor activity of UPS inhibition to solid malignancies [46].

5   Simultaneous Targeting of DUBs and Autophagy

Targeting components of the autophage–lysosome system and the ubiquitin–proteasome system as anticancer therapy have shown initial promise but have also been plagued by problems of resistance and/or toxicity. Accumulating evidence has supported a cross talk between these two protein degradation pathways, where one system can compensate for a weakness in the other if activity is hindered [47]. This ability to adapt to inhibition of either autophagy or the UPS could be a large contributor to the resistance observed in clinics. Recent understanding of this cross talk triggered interest in simultaneous inhibition of the UPS and autophagy as a new strategy to synergistically induce cell death in multiple cancer settings. By blocking both protein degradation systems at once, the cell will lose its ability to compensate and overcome inhibition of a single system and therefore succumb to cell death (Fig. 1).

Encouragingly, combination of DUB and autophagy inhibition has shown antitumor effects against solid tumors. Small-molecule DUB inhibitors such as b-AP15 and RA-9 have a profound effect on triple-negative breast cancer (TNBC) viability and have led to activation of autophagy as a cellular mechanism to compensate for ubiquitin–proteasome-system stress. Importantly, the activation of autophagy occurred well before induction of apoptosis, which indicated the protective and pro-survival nature of its induction. Together this supported rationale for concomitant inhibition of DUBs and autophagy in the TNBC model. To determine whether or not inhibitors would have a synergistic effect, TNBC cells were treated with suboptimal doses of DUB and lysosome inhibitors. Although the doses would be suboptimal on their own, the combination of inhibitors caused synergistic cell death of TNBC cells [48]. Accordingly, the evaluation of DUB inhibition in combination with lysosomal inhibition should be further explored as a therapeutic approach for the treatment of TNBC and other malignancies.
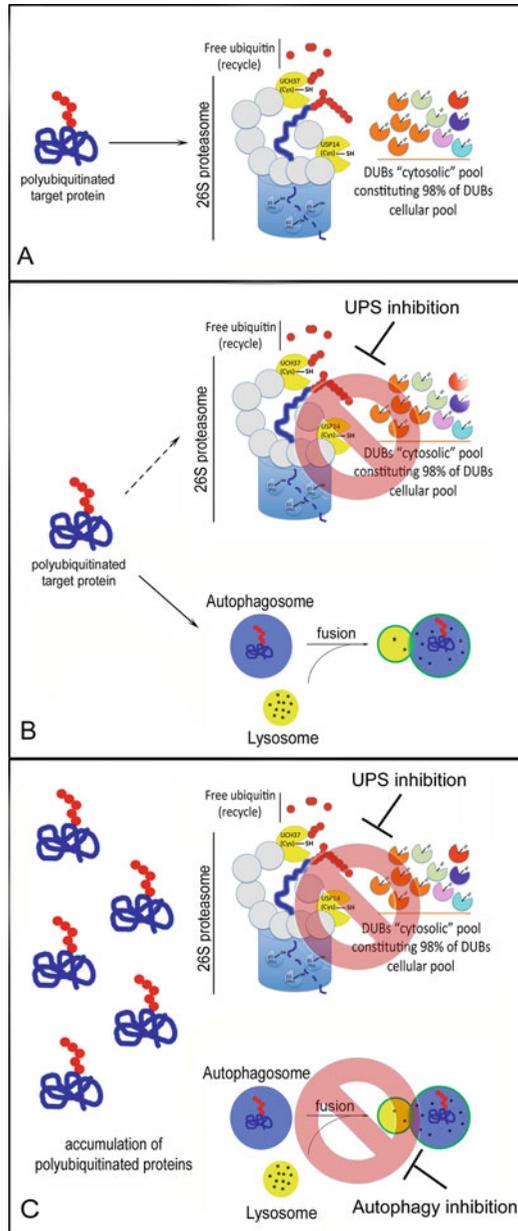
Fig. 1 Synergistic approach to overcome cross talk between UPS and autophagy. (**a**) Conventional pathway for 26S proteasome degradation of polyubiquitin-tagged target proteins by the ubiquitin–proteasome system (UPS). Ubiquitin tags are removed by deubiquitinating enzymes (DUBs) before the target protein is threaded through the 26S proteasome for degradation. (**b**) The autophage–lysosome system can compensate for degradation of ubiquitin-tagged proteins when UPS function is compromised. The red cancel icon is superimposed over both the DUBs and the proteasome to reflect the various methods available to inhibit the UPS, such as proteasomal inhibition or DUB inhibition. (**c**) Simultaneous targeting of the UPS and autophagy results in an accumulation of polyubiquitinated proteins within a cell. Red cancel icons indicate inhibition of both protein degradation systems. The placement of the red cancel icon over the autophage–lysosome system encompasses different targets of autophagy inhibition, such as fusion of autophagosomes and lysosomes or de-acidification of the lysosome

6   Conclusion and Outlook

The ubiquitin–proteasome system and the autophage–lysosome system are two main protein degradation pathways within a eukaryotic cell that work to maintain proper protein homeostasis. These pathways are heavily relied upon by cancer cells due to their increased burden of misfolded proteins resulting from rapid proliferation rates, as well as their need for metabolic resources in nutrient-poor environments such as a tumor. Due to these factors, cancer therapeutics based on inhibition of both the autophage–lysosome pathway and the ubiquitin proteasome pathways have been explored. Although this field has experienced remarkable successes, such as the implementation of bortezomib in multiple myeloma treatment regimens, the success has been limited by side effects and common development of resistance.

Cross talk between the autophage–lysosome and ubiquitin–proteasome pathways has recently been suggested in cancer, with evidence signifying they are functionally coupled. This provides a possible explanation for why resistance is so prominent in response to these treatments: as one protein degradation pathway is inhibited, the other can compensate. Recent studies have suggested simultaneous targeting of both protein degradation pathways to overcome the issue of cross talk. Specifically, interest has been taken in targeting the deubiquitinating enzymes of the UPS. With almost 100 DUBs identified, deubiquitinating enzymes have the potential to provide previously unachievable levels of specificity targeting each cancer type when compared to proteasome inhibition. Using a combinational approach of targeting DUBs and autophagy could profoundly increase efficacy and duration of response to protein degradation inhibitors and translate their anticancer properties to variety of cancer settings previously unresponsive to a single-agent approach.

References

1. Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. Nature 426:895–899
2. Amm I, Sommer T, Wolf DH (2014) Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. Biochim Biophys Acta 1843:182–196
3. Levine B (2007) Cell biology: autophagy and cancer. Nature 446:745–747
4. Hoeller D, Dikic I (2009) Targeting the ubiquitin system in cancer therapy. Nature 458:438–444
5. Yang ZJ, Chee CE, Huang S, Sinicrope FA (2011) The role of autophagy in cancer: therapeutic implications. Mol Cancer Ther 10:1533–1541
6. Frezza M, Schmitt S, Dou QP (2011) Targeting the ubiquitin-proteasome pathway: an emerging concept in cancer therapy. Curr Top Med Chem 11:2888–2905
7. Boya P, Reggiori F, Codogno P (2013) Emerging regulation and functions of autophagy. Nat Cell Biol 15:713–720
8. Thorburn A, Thamm DH, Gustafson DL (2014) Autophagy and cancer therapy. Mol Pharmacol 85:830–838
9. Mizushima N (2007) Autophagy: process and function. Genes Dev 21:2861–2873

10. Lin Z, Bazzaro M, Wang MC et al (2009) Combination of proteasome and HDAC inhibitors for uterine cervical cancer treatment. Clin Cancer Res 15:570–577

11. Bazzaro M, Lin Z, Santillan A et al (2008) Ubiquitin proteasome system stress underlies synergistic killing of ovarian cancer cells by bortezomib and a novel HDAC6 inhibitor. Clin Cancer Res 14:7340–7347

12. Yang Z, Klionsky DJ (2010) Eaten alive: a history of macroautophagy. Nat Cell Biol 12:814–822

13. Liang XH, Jackson S, Seaman M et al (1999) Induction of autophagy and inhibition of tumorigenesis by beclin 1. Nature 402: 672–676

14. Mathew R, Karantza-Wadsworth V, White E (2007) Role of autophagy in cancer. Nat Rev Cancer 7:961–967

15. Kung CP, Budina A, Balaburski G et al (2011) Autophagy in tumor suppression and cancer therapy. Crit Rev Eukaryot Gene Expr 21:71–100

16. Guo JY, Xia B, White E (2013) Autophagy-mediated tumor promotion. Cell 155:1216–1219

17. Dolfi SC, Chan LL, Qiu J et al (2013) The metabolic demands of cancer cells are coupled to their size and protein synthesis rates. Cancer Metab 1:20

18. Kim SE, Park HJ, Jeong HK et al (2015) Autophagy sustains the survival of human pancreatic cancer PANC-1 cells under extreme nutrient deprivation conditions. Biochem Biophys Res Commun 463:205–210

19. Zhi X, Zhong Q (2015) Autophagy in cancer. F1000Prime Rep 7:18

20. Zhang Y, Liao Z, Zhang LJ, Xiao HT (2015) The utility of chloroquine in cancer therapy. Curr Med Res Opin 31:1009–1013

21. Lee JY, Koga H, Kawaguchi Y et al (2010) HDAC6 controls autophagosome maturation essential for ubiquitin-selective quality-control autophagy. EMBO J 29: 969–980

22. Nagelkerke A, Bussink J, Geurts-Moespot A et al (2015) Therapeutic targeting of autophagy in cancer. Part II: Pharmacological modulation of treatment-induced autophagy. Semin Cancer Biol 31:99–105

23. Ozpolat B, Benbrook DM (2015) Targeting autophagy in cancer management—strategies and developments. Cancer Manag Res 7: 291–299

24. Johnson DE (2015) The ubiquitin-proteasome system: opportunities for therapeutic intervention in solid tumors. Endocr Relat Cancer 22:T1–T17

25. Rajkumar SV, Richardson PG, Hideshima T, Anderson KC (2005) Proteasome inhibition as a novel therapeutic target in human cancer. J Clin Oncol 23:630–639

26. Chen D, Frezza M, Schmitt S et al (2011) Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets 11: 239–253

27. Frankland-Searby S, Bhaumik SR (2012) The 26S proteasome complex: an attractive target for cancer therapy. Biochim Biophys Acta 1825:64–76

28. Kane RC, Bross PF, Farrell AT, Pazdur R (2003) Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. Oncologist 8:508–513

29. Dick LR, Fleming PE (2010) Building on bortezomib: second-generation proteasome inhibitors as anti-cancer therapy. Drug Discov Today 15:243–249

30. Papadopoulos KP, Burris HA 3rd, Gordon M et al (2013) A phase I/II study of carfilzomib 2-10-min infusion in patients with advanced solid tumors. Cancer Chemother Pharmacol 72:861–868

31. Dou QP, Zonder JA (2014) Overview of proteasome inhibitor-based anti-cancer therapies: perspective on bortezomib and second generation proteasome inhibitors versus future generation inhibitors of ubiquitin-proteasome system. Curr Cancer Drug Targets 14:517–536

32. Hussain S, Zhang Y, Galardy PJ (2009) DUBs and cancer: the role of deubiquitinating enzymes as oncogenes, non-oncogenes and tumor suppressors. Cell Cycle 8:1688–1697

33. Lim KH, Baek KH (2013) Deubiquitinating enzymes as therapeutic targets in cancer. Curr Pharm Des 19:4039–4052

34. Reyes-Turcu FE, Ventii KH, Wilkinson KD (2009) Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. Annu Rev Biochem 78:363–397

35. Amerik AY, Hochstrasser M (2004) Mechanism and function of deubiquitinating enzymes. Biochim Biophys Acta 1695:189–207

36. Ndubaku C, Tsui V (2015) Inhibiting the deubiquitinating enzymes (DUBs). J Med Chem 58:1581–1595

37. Yang JM (2007) Emerging roles of deubiquitinating enzymes in human cancer. Acta Pharmacol Sin 28:1325–1330

38. McClurg UL, Robson CN (2015) Deubiquitinating enzymes as oncotargets. Oncotarget 6:9657–9668

39. D'Arcy P, Wang X, Linder S (2015) Deubiquitinase inhibition as a cancer therapeutic strategy. Pharmacol Ther 147:32–54

40. Anchoori RK, Karanam B, Peng S et al (2013) A bis-benzylidine piperidone targeting proteasome ubiquitin receptor RPN13/ADRM1 as a therapy for cancer. Cancer Cell 24:791–805

41. Kapuria V, Peterson LF, Fang D et al (2010) Deubiquitinase inhibition by small-molecule WP1130 triggers aggresome formation and tumor cell apoptosis. Cancer Res 70:9265–9276

42. Donato NJ, Talpaz M, Peterson L et al (2015) Deubiquitinase inhibitors and methods for use of the same. Google Patents Publication number WO2015054555 A1

43. Tian Z, D'Arcy P, Wang X et al (2014) A novel small molecule inhibitor of deubiquitylating enzyme USP14 and UCHL5 induces apoptosis in multiple myeloma and overcomes bortezomib resistance. Blood 123:706–716

44. Wang X, D'Arcy P, Caulfield TR et al (2015) Synthesis and evaluation of derivatives of the proteasome deubiquitinase inhibitor b-AP15. Chem Biol Drug Des 86:1036–1048

45. Coughlin K, Anchoori R, Iizuka Y et al (2014) Small-molecule RA-9 inhibits proteasome-associated DUBs and ovarian cancer in vitro and in vivo via exacerbating unfolded protein responses. Clin Cancer Res 20:3174–3186

46. Kushwaha D, O'Leary C, Cron KR et al (2015) USP9X inhibition promotes radiation-induced apoptosis in non-small cell lung cancer cells expressing mid-to-high MCL1. Cancer Biol Ther 16:392–401

47. Korolchuk VI, Menzies FM, Rubinsztein DC (2010) Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. FEBS Lett 584:1393–1398

48. Vogel RI, Coughlin K, Scotti A et al (2015) Simultaneous inhibition of deubiquitinating enzymes (DUBs) and autophagy synergistically kills breast cancer cells. Oncotarget 6:4159–4170

Chapter 6

Quantitative Clinical Imaging Methods for Monitoring Intratumoral Evolution

Joo Yeun Kim and Robert A. Gatenby

Abstract

Solid tumors are multiscale, open, complex, dynamic systems: *complex* because they have many interacting components, *dynamic* because both the components and their interactions can change with time, and *open* because the tumor freely communicates with surrounding and even distant host tissue. Thus, it is not surprising that striking intratumoral variations are commonly observed in clinical imaging such as MRI and CT and that several recent studies found striking regional variations in the molecular properties of cancer cells from the same tumor. Interestingly, this spatial heterogeneity in molecular properties of tumor cells is typically ascribed to branching clonal evolution due to accumulating mutations while macroscopic variations observed in, for example, clinical MRI scans are usually viewed as functions of blood flow. The clinical significance of spatial heterogeneity has not been fully determined but there is a general consensus that the varying intratumoral landscape along with patient factors such as age, morbidity and lifestyle, contributes significantly to the often unpredictable response of individual patients within a disease cohort treated with the same standard-of-care therapy.

Here we investigate the potential link between macroscopic tumor heterogeneity observed by clinical imaging and spatial variations in the observed molecular properties of cancer cells. We build on techniques developed in landscape ecology to link regional variations in the distribution of species with local environmental conditions that define their habitat. That is, we view each region of the tumor as a local ecosystem consisting of environmental conditions such as access to nutrients, oxygen, and means of waste clearance related to blood flow and the local population of tumor cells that both adapt to these conditions and, to some extent, change them through, for example, production of angiogenic factors. Furthermore, interactions among neighboring habitats can produce broader regional dynamics so that the internal diversity of tumors is the net result of complex multiscale somatic Darwinian interactions.

Methods in landscape ecology harness Darwinian dynamics to link the environmental properties of a given region to the local populations which are assumed to represent maximally fit phenotypes within those conditions. Consider a common task of a landscape ecologist: defining the spatial distribution of species in a large region, e.g., in a satellite image. Clearly the most accurate approach requires a meter by meter survey of the multiple square kilometers in the region of interest. However, this is both impractical and potentially destructive. Instead, landscape ecology breaks the task into component parts relying on the Darwinian interdependence of environmental properties and fitness of specific species' phenotypic and genotypic properties. First, the satellite map is carefully analyzed to define the number and distribution of habitats. Then the species distribution in a representative sampling of each habitat is empirically determined. Ultimately, this permits sufficient bridging of spatial scales to accurately predict spatial distribution of plant and animal species within large regions.

Currently, identifying intratumoral subpopulations requires detailed histological and molecular studies that are expensive and time consuming. Furthermore, this method is subject to sampling bias, is invasive for

vital organs such as the brain, and inherently destructive precluding repeated assessments for monitoring post-treatment response and proteogenomic evolution. In contrast, modern cross-sectional imaging can interrogate the entire tumor noninvasively, allowing repeated analysis without disrupting the region of interest. In particular, magnetic resonance imaging (MRI) provides exceptional spatial resolution and generates signals that are unique to the molecular constituents of tissue. Here we propose that MRI scans may be the equivalent of satellite images in landscape ecology and, with appropriate application of Darwinian first principles and sophisticated image analytic methods, can be used to estimate regional variations in the molecular properties of cancer cells.

We have initially examined this technique in glioblastoma, a malignant brain neoplasm which is morphologically complex and notorious for a fast progression from diagnosis to recurrence and death, making a suitable subject of noninvasive, rapidly repeated assessment of intratumoral evolution. Quantitative imaging analysis of routine clinical MRIs from glioblastoma has identified macroscopic morphologic characteristics which correlate with proteogenomics and prognosis. The key to the accurate detection and forecasting of intratumoral evolution using quantitative imaging analysis is likely to be in the understanding of the synergistic interactions between observable intratumoral subregions and the resulting tumor behavior.

Key words Diagnostic imaging methods, Ecology, Genetic predisposition to disease, Individualized medicine, Oncology, Darwinian dynamics, Intratumoral heterogeneity, Evolutionary biology, Proteogenomics, Treatment resistance and disease recurrence

1   Introduction

Patients who share a common oncologic diagnosis and are treated with identical standard-of-care therapy often display a wide range of responses and survivorship. Some of this variation is likely due to patient characteristics such as age and comorbidities as well as interpatient molecular heterogeneity resulting in, for example, differences in drug metabolism and toxicity. However, uniquely among human diseases, cancer cells can evolve and adapt to overcome threats and take advantage of opportunities. Thus, while it is not surprising that tumor cells vary between patients, local Darwinian dynamics characteristically produce molecular variations between tumors in the same patient and even in different malignant cells in the same tumor [1–5]. This heterogeneity has significant clinical consequences. Consider, for example, recent advances in tumor treatment that use small molecules that target particular molecular pathways that promote proliferation. For example, cetuximab targets the epidermal growth factor receptor (EGFR) in lung cancer and head and neck cancer. By identifying the EGFR target on the tumor cells prior to therapy, the probability that cetuximab will produce a positive clinical response is greatly increased. However, the effects of cetuximab are almost invariably time-limited because the tumor cells evolve resistance and, importantly, the resistant cells are generally thought to be present in the tumor prior to therapy due to intratumoral heterogeneity. Such resistance leading to subsequent progression is common particularly for system treatment, and additional biomarkers are rarely available for second, third, and fourth line therapies. The success of Darwinian dynamics in overcoming cancer therapy is hardly unique to targeted treatments. In

fact, the vast majority of disseminated cancers remain fatal despite the development of many new therapies that initially produce significant, often dramatic, results. In almost every case, cancer cells use acquired mutations or simply access and upregulate information from the vast information storehouse of the human genome, to defeat the therapy and permit tumor recurrence. In other words, evolution is the proximate cause of death in most patients treated for cancer, and this will likely remain so even with new advances in targeted therapy and immunotherapy.

Thus, there is a pressing need to develop methods to evaluate intratumoral evolutionary dynamics during therapy. However, achieving this requires solutions to a number of challenges posed by tumor heterogeneity, particularly spatial variations of intratumoral population dynamics. A number of studies have demonstrated substantial molecular variation in different tumors and even in different regions of the same tumor [1–5]. Although a single therapy may be applied to every region of the tumor, it is acting on regionally different populations which, in turn, will result in variable levels of response and resistance.

Currently, regional variations in tumor cell molecular properties and changes in that distribution over time during therapy can only be established by multiple repeated biopsies. But this is limited by cost, patient compliance, and potential complications. A few less invasive mechanisms to estimate intratumoral evolution are available. For example, serial sampling of tumor cells or molecules in the blood samples may reflect the global evolutionary dynamics within the total patient tumor burden. We have proposed that imaging, because it can repeatedly, noninvasively, and nondestructively characterize spatial variations within tumors, has the potential to serve as a biomarker of intratumoral evolution during therapy. In particular, we have focused on MRI scans, which have no ionizing radiation and can repeatedly interrogate the tumor with multiple sequences that are sensitive to different components of the tumor ecology such as blood flow, interstitial edema, and cellular density.

Radiologists have long been aware of significant and consistent spatial heterogeneity in the imaging characteristics of tumors. Furthermore, these variations often change with time during therapy or tumor progression. This leads to a central question: Are the spatial changes observed in clinical imaging directly linked to regional variations in the cellular, microenvironmental, and molecular properties observed within and between tumors?

Answering this question requires solving multiple challenges inherent in comparing data at very different spatial and temporal scales. Interestingly, this is highly analogous to problems in landscape ecology, which studies the distribution and abundance of species in relation to the scale (fine- versus coarse-grained heterogeneity) and proportions of different habitats within an ecosystem. The properties of the habitat can be used to infer the dominant local species. For example, in Fig. 1 [6], the spatial variations in the
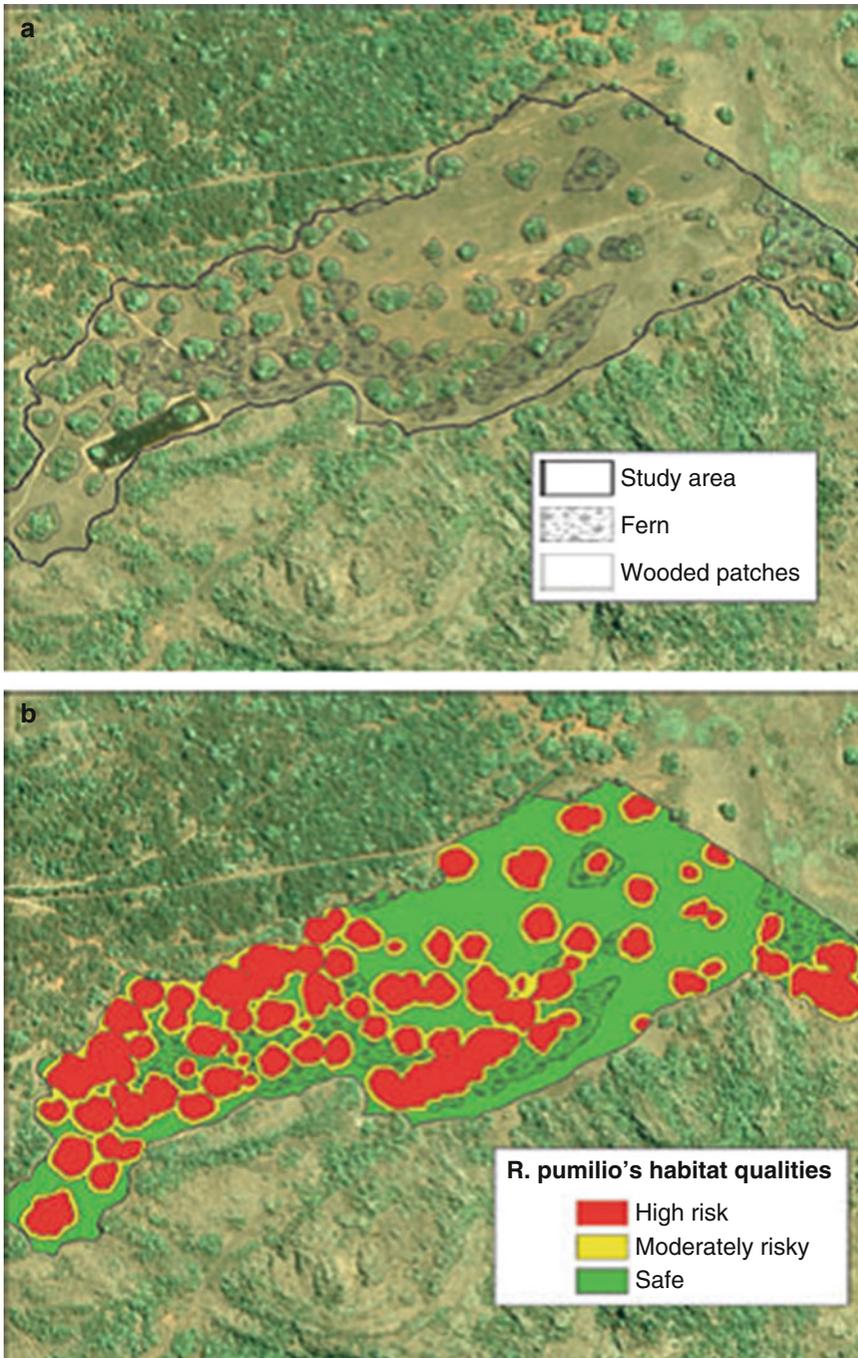
Fig. 1 Satellite imagery used to form landscape ecology maps. (**a**) The study area showing polygons around the main habitats of *Rhabdomys pumilio*. (**b**) The study area showing habitat classification according to *Rhabdomys pumilio*'s perceived risk of predation. Landscape ecology maps use selective samples of specific habitats to bridge spatial scales and estimate species distribution in a large area containing multiple habitats. Reproduced from Baker and Brown, 2010 (Ref. 6) with permission from John Wiley & Sons

density of chipmunks (*Rhabdomys pumilio*) can be predicted by identifying the habitat characteristics observable on satellite images based on understanding of the features that favor the presence or absence of predators.

To illustrate this approach, we focus on glioblastoma (GBM), which is the most common adult brain neoplasm and uniformly fatal. Upon tissue diagnosis, standard of care (SOC) treatment usually consists of surgical resection or debulking followed by combined radiation therapy and temozolomide, an orally administered alkylating agent. Despite this vigorous, multimodal therapy, tumor progression is virtually inevitable and overall survival is usually 12–14 months [7] following diagnosis. However, there is significant variation within this cohort—some die within 2 months of diagnosis while a small percentage experiences an unusually long term (>4 years) survival.

The varying clinical outcomes following SOC therapy suggest that resistant cancer cells exist prior to therapy and/or develop during treatment. Identifying and targeting these resistant populations would likely improve outcomes but how can this be achieved? Interrogation of the heterogeneity of glioblastoma during therapy through repeated tissue/cellular/molecular analysis seems to be the obvious answer. However, this is impractical based on cost and compliance and may be dangerous as repeated biopsies of the brain lead to loss of memory, personality, and function [8] as well as cost of money and time. Furthermore, the sampling process can alter the tumor dynamics, for example, through hemorrhage.

Thus, while the wide range of sensitivity and resistance to treatment that is found in GBM, and many other malignancies, is likely to be at least partly attributable to intratumoral proteogenomic heterogeneity, the clinical characterization and investigation of the spatial variations in macro- or micro-scale tumor properties is difficult so that the significance of observed molecular variations [9] remains largely unknown.

To address these issues we note that sophisticated cross-sectional diagnostic imaging modalities including magnetic resonance imaging (MRI) and computed tomography (CT) can noninvasively and nondestructively characterize tissue-level intratumoral variations in GBMs repeatedly over time during and after therapy. In a number of studies, we and others have investigated the potential of imaging-defined tissue heterogeneity to provide insights in spatial and temporal variations in the cellular and molecular characteristics of tumor cells. We hypothesize that the tools to bridge the spatial scales may exist within landscape ecology as described above. Similar to satellite images, MRI and CT scans can define relatively large scale structures within a tumor. It is possible that imaging characteristics can be used to define regional variations in the ecology of each cancer which, in turn, will correlate with a "species map" of spatial distribution of distinct phenotypic and genotypic cancer subpopulations.

2   Quantitative Methods in Clinical Cancer Imaging

An initial challenge for this work is development of quantitative and reproducible metrics to characterize the imaging properties of cancers. Medical imaging is an integral part of oncologic diagnosis and post-treatment surveillance but monitoring strategies such as the Response Evaluation Criteria in Solid Tumors (RECIST) and World Health Organization (WHO) criteria focus only on utilizing the tumor size as the metric of comparison. Inter-observer variability is a common and well-known short coming in all of these methods. The documentation of the perceivable intratumoral textural variations, necrosis, enhancement patterns and interstitial edema remains descriptive and subject to large inter-observer variability. Because morphologic and textural assessment is limited to the macroscopic spatial resolution, even the most experienced eye can fail to recognize a small pathologic nodule or lymph node which may appear morphologically normal but be proven to be malignant through functional imaging such as positron emission tomography (PET). Similarly, the eye can fail to appreciate the fine granular variations in imaging signal characteristics and texture (Fig. 2) [9] that, if detected, could lead to a different diagnosis, or a different cancer stage.

3   Radiomics

While each CT or MRI requires several megabits of storage, none of this data is typically quantitative. Furthermore, as noted earlier, interpretation is typically reliant entirely on visual inspection. In fact, cross-sectional imaging modalities are potential sources of a huge volume of mineable data. Furthermore, MR images, for example, are built from single pixels each displaying the unique echo signal generated by the molecular constituents of the corresponding minute volume of tissue. A close look at an apparently uniform region of tumor can reveal subtle spatial variations of signals which likely reflect the underlying cellular and molecular heterogeneity (Fig. 3). An automated method of quantifying these signal variations that are visually imperceptible may provide an accurate and objective metric for assessing a patient's prognosis. For example, Lacroix et al. [10] found that quantitation of necrosis and enhancement on preoperative MRI scans were independent predictors of GBM survival. Several other studies [2, 10–13] have shown that MRI characteristics (e.g., ratio of tumor enhancement to necrosis) are predictive of survival. Thus, a quantified measurement of the necrosis and enhancement prior to treatment may provide a further dimension to guide treatment planning, and be compared with subsequent post-treatment measurements as an objective means of monitoring treatment response.
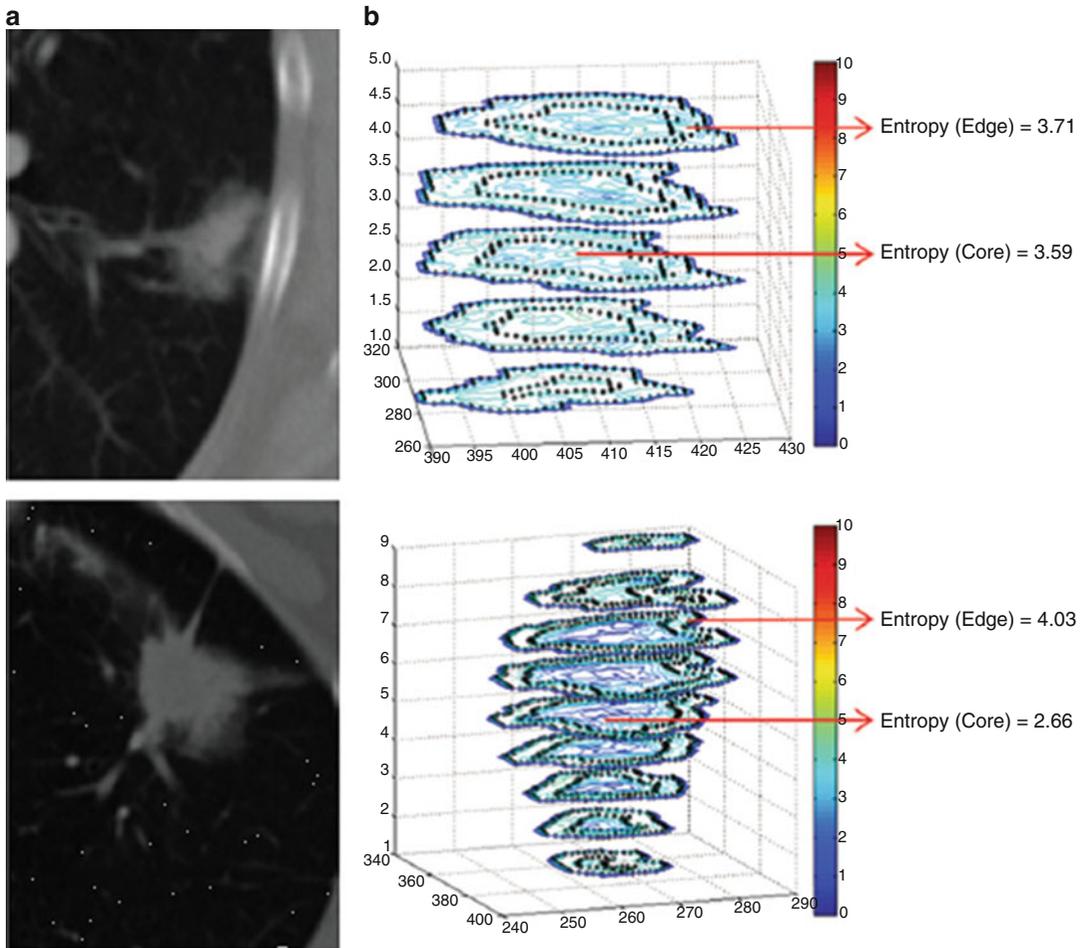
Fig. 2 (**a**) CT images obtained with conventional entropy filtering in two patients with non-small cell lung cancer with no apparent textural differences show similar entropy values across all sections. (**b**) Contour plots obtained after the CT scans were convolved with the entropy filter. Further subdividing each section in the tumor stack into tumor edge and core regions (*dotted black contour*) reveals varying textural behavior across sections. Two distinct patterns have emerged, and preliminary analysis shows that the change of mean entropy value between core and edge regions correlates negatively with survival. Reproduced from Gatenby et al., 2013 with permission from RSNA (Ref. 9)

The general approach to develop quantitative methods and potential data mining algorithms from conventional clinical images is termed "radiomics." These methods allow automated high throughput imaging feature extraction from radiologic images and offers a tool for fine-grained, operator-independent diagnostic and follow-up methods that are objectively quantitative, reproducible, and repeatable for monitoring temporal intratumoral evolution.

Importantly, however, virtually all radiomic techniques view the tumor as a single entity and thus produce metrics that are not spatially explicit. In the remainder of this chapter, we focus on efforts to develop methods that use imaging to define regional

T1post-T1pre	FLAIR	Color
High	High	
Low	High	
High	Low	
Low	Low	
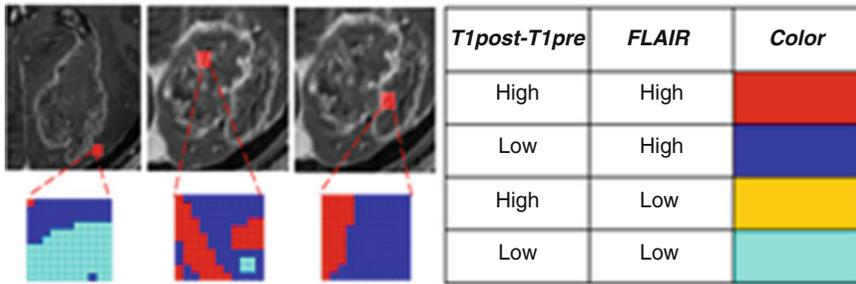
Fig. 3 Patient from our study cohort showing three biopsy sites (*red boxes* in the MRIs) from within the tumor core (all designated as contrast-enhancing region CE). While macroscopic inspection shows similar characteristics of the peritumoral region, multiparametric analysis shows that these are in fact different (Color figure online)

variations in the tumor and initial attempts to use ecological principles to link these macroscopic scale features to the molecular and cellular properties of local tumor populations.

4   The Habitat Concept of Intratumoral Heterogeneity

Many tumor types have been found to exhibit considerable intratumoral variation through RNA–seq studies and other molecular and cellular analysis [8, 14–19]. A single glioblastoma tumor will harbor multiple cellular subtype classifiers, along with heterogeneity in transcription programs related to oncologic signaling, proliferation, immune response, and hypoxia [20]. No two tumors are alike, but what determines their unique intratumoral organization? Such spatially explicit distribution of species are seen in landscape ecological studies which is built upon Darwinian dynamics in which evolution is governed by interactions of an organism's phenotype with local environmental properties that act as selection forces. Thus, geographic distributions or species are not the result of some random mutational process but identifiable and predictable interactions with the varying landscape and microenvironments. In cancer, initial tumor growth is likely to be determined by the accumulation of mutations and resultant aberrations of cell cycling. However, as the tumor increases in size, regional variations usually emerge under the influence of its immediate anatomic constraints, its proximity to blood vessels and the patient's immunological response. Thus, the dynamics that drive somatic evolution within cancers, as in nature, are not simply governed by accumulating mutations but rather, by interactions among local environmental selection forces and regional phenotypic adaptations to those forces.

In other words, the molecular properties of cancer cells are the "how" of evolution but the local environmental conditions are the "why." This linkage is the basis for our proposal that conventional

clinical imaging can be used to predict intratumoral molecular heterogeneity. That is, intratumoral Darwinian dynamics fundamentally connect the intratumoral and environmental conditions such as blood flow, interstitial edema, and cellular density. The latter properties are readily identifiable on MRI scans. This leads to the hypothesis that identifiable and quantifiable MRI-defined "habitats" can predict regional distribution of cellular and molecular properties observable on pathology.

5 Methods in Landscape Ecology

One of the challenges in the analysis of intratumoral heterogeneities is in deciding how to measure the spatial variations. Similar questions are answered in landscape ecology using spatial statistical methods. Landscape ecology studies the distribution and abundance of species in relation to the scale (fine- versus coarse-grained heterogeneity) and proportions of distinct habitats within an ecosystem. It is inherently scale dependent and relies on concepts from population dynamics and niche theory, and on analyses associated with spatial and temporal autocorrelations, cluster analyses, and structured equation modeling to tease apart direct and indirect effects within causal relationships [21]. For example, when tasked with measuring plant species distributions in a large heterogeneous region such as the Smokey Mountain National Park, landscape ecologists typically begin by reviewing satellite imagery of the entire region (Fig. 1) [6] to identify and analyze distinct habitats via pixilation of variables such as color, reflectance, infrared and UV intensities.

Such analyses can be used to discriminate both the number and character of distinct habitats. In lockstep with the satellite image, the different habitats are "ground-truthed" by more detailed, small-scale sampling and mapping of "quadrats" that are stratified by the habitats noted from the satellite image. This quantifies the species composition of each habitat, and cross-checks the heterogeneity from the quadrats with those from the satellite images. By relating small-scale sampling with larger scale imaging, it is now possible to map the distribution of species. Such strategies are being used to monitor the speed of rainforest destruction, and estimate the total number of an endangered species.

On a smaller order of magnitude, individual habitats within a microcosm can be identified by the number of animals with distinct niches. The number of habitats within a microcosm can be calculated using the K-factor clustering technique, usually dependent on the number of inhabitant species. A numerical representation of a relationship between a habitat and its desirability to a species is the Habitat Suitability Index expressed as a score between 0 and 1, 0 being undesirable and 1 being more desirable.

Establishing the geographic distribution of different species usually requires identification of adaptational behavioral variations between species that occupy similar niches, for example between the grey and fox squirrels. The grey squirrels are excellent foragers but lack survival strategies when they encounter predators. Thus, they uniformly and predictably inhabit wooded areas around human habitats where predators are rare and survival is largely dependent of acquisition of nutrients. While they are excellent scavengers, they seldom encounter predators thus show a relative lack of survival strategies when faced by such danger. In comparison, the fox squirrels, while less effective in foraging for nutrients, fiercely and effectively defend themselves against predators. Thus, fox squirrels are predictably found to be dominant in poorly wooded areas on the outskirts of cities. Thus, by defining the habitat characteristics on satellite maps, the distribution of squirrel species can be reliably estimated.

We hypothesize that intratumoral heterogeneity visible on MRI can be viewed as a collection of "habitats" each with a distinct ecosystem. Using principles and techniques from landscape ecology we propose that the coarse-grained, intratumoral variations in clinical imaging could be defined by the local environmental selection forces, and could predict the distribution of fine-grained features such as cellular and molecular properties.

6   Evidence of Intratumoral Habitats in Medical Imaging

Like satellite images, the intratumoral and peritumoral regional variations are well demonstrated by Magnetic resonance imaging (MRI) through the use of sequences sensitive to blood flow, vascular permeability, interstitial edema, and cell density [2, 11–13, 20]. Aggressive tumors such as glioblastoma are radiologically recognized by their rapid growth, high vascularity, penchant for building an unstable blood supply often leading to central necrosis and cavitation, and a high volume of local interstitial edema. The ability of medical imaging to display the intratumoral and intertumoral spatial heterogeneity in glioblastoma is well recognized [1, 7, 17, 18] (Fig. 4).

In our view, the presence of such spatial heterogeneity indicates that each tumor cannot be viewed as a single entity. Instead, tumors may be seen to be an ecosystem of distinct but dynamically interacting habitats that can be seen as multiscale ecologies adapting to, and in turn affecting, the selection forces in its environment.

To identify the inhabitants of such habitats, investigators have examined the "radio-genomics" and the "radio-pathomics" in glioblastoma. In particular, Barajas et al. [1] took MRI-guided biopsies from the contrast-enhancing (designated CE) and non-enhancing infiltrative (NE) tumor regions and found that the CE had increased tumor score, cellular density, proliferation and architectural disruption in comparison to NE. Gutman et al. [2] found
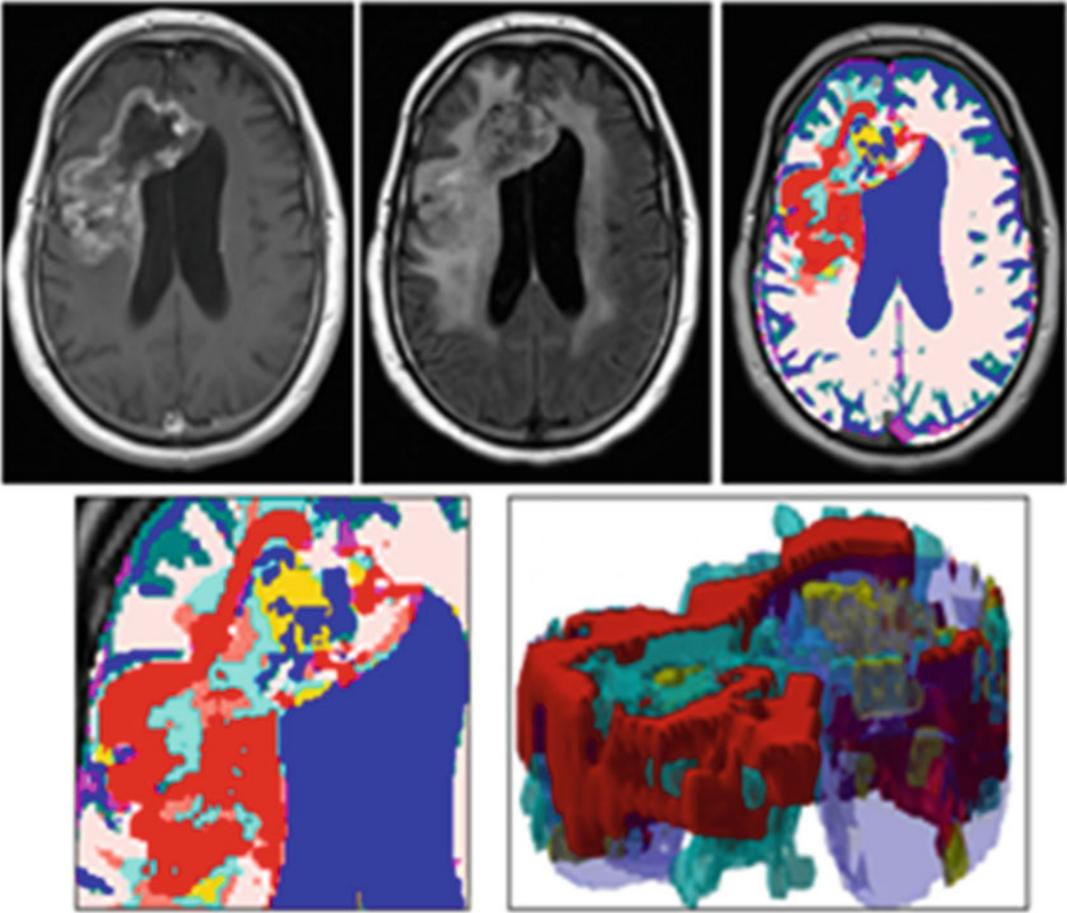
Fig. 4 Whole brain and close-up images of a recurrent GBM tumor. The *top row* shows T1 post-gadolinium contrast (T1Gd), Fluid-attenuated Inversion Recovery (FLAIR) and a 3D multiparametric tumor cluster map constructed from superimposition of T1 pre- and post-gadolinium contrast, T2, FLAIR, and apparent diffusion coefficient (ADC) sequences. *Each color* represents a particular combination of multiparametric signals, or "habitat," defined by the analytic methods described in the text, and demonstrates the spatial complexity of MRI data (Color figure online)

that low-enhancing areas of a glioblastoma would be dominated by the proneural subtype, and the highly enhancing regions be inhabited by the mesenchymal subtype. Increased mass effect as well as a number of distinct imaging features correlated with increased EGFR expression and intratumoral gene-expression patterns [11]. Similarly, several prior studies examined global MRI features such as the infiltrative pattern, enhancement, edema, and necrosis as prognostic biomarkers [2, 3, 5, 11–13, 22–29]. Lev et al. [3] used a decision tree technique to identify prognostic imaging features. Gutman et al. [2] correlated tumor size and related structural composition with clinical survival. A number of studies have also identified intratumoral and intertumoral variations in the molecular properties of glioblastoma cells to have clinical significance [3–5].

These studies demonstrate that tumor imaging features can be predictive of the underlying proteogenomics and prognosis, thereby provide a way to noninvasively predict tumor regions which have the potential for treatment failure. Harnessing the MRI's ability to accurately demonstrate the environmental components, in particular vascularity, may lead to the identification of treatment resistant areas of a tumor. The most important environmental influence on the growth of a tumor is the blood supply which is directly related to local concentrations of oxygen, glucose, serum growth factors, and acid clearance. The outer edges of a tumor have easy access to blood supply, providing an oxygen- and nutrient-rich environment for tumor cells which, are also under constant attack from cell mediated immunity. Tumor cells at the edge of the tumor are "pioneer-like," inducing angiogenesis, locally infiltrating, and withstanding the harsh predator-like dynamics imposed by immune response.

With an increasing tumor volume, the central regions of the tumor become poorly perfused and increasingly hypoxic and nutrient poor. Some tumor cells enter apoptosis resulting in necrosis, however some can adopt the anaerobic respiration, producing acidic by-products of glucose metabolism. A poor blood supply also means impeded waste disposal, making these tumor regions perpetually acidic [7, 9, 30–34]. Tumor cells surviving in these regions can be considered as "engineers," with adaptive mechanisms that enable them to endure the acidic and hypoxic environment, with slowed cell cycling kinetics, being sparse in number but apoptosis resistant. Such cells sustain a stable population and display treatment resistance through poor chemotherapeutic agent delivery and poor radiosensitivity [35, 36]. On imaging, such tumor habitats can be expected to show a low level of contrast enhancement and maintenance of a reasonable cellular density.

This intratumoral variation presents a significant obstacle on the road to "personalized" cancer therapy, which is currently based on a small number of tumor biopsy samples assumed to be representative of a tumor's molecular properties [37–39], and imaging features that average the geographical intratumoral heterogeneity.

Thus, identifying discrete imaging features that consistently predict targetable pathologic tumor characteristics such as vascular density and driver gene expression would be of significant clinical value.

7 A Practical Approach to Radiomics and Quantitative Habitat Imaging on Magnetic Resonance Imaging

Through the collaboration of evolutionary biologists, imaging scientists and radiologists, we have divided glioblastoma tumors into "habitats" based on MRI multiparametric characteristics. Central to this investigation is our conceptual model of glioblastoma as

interacting subregions or habitats each with a distinct ecology defined by Darwinian selection forces from the local environmental properties (usually governed by blood flow influencing local concentrations of oxygen, glucose, acid, and serum growth factors) and cellular phenotypic characteristics that are adapted to these conditions [7, 9, 30–34, 40]. Since different MRI sequences are sensitive to blood flow, interstitial edema, and cell density, we reason that MRI can define key components of the intratumoral habitats.

To analyze the multiscale ecology of a glioblastoma and its clinical significance, we have developed and applied innovative methodologies including new image analytic methods to generate novel three-dimensional multiparametric MRI maps of GBM (Fig. 5) [9] which can be used to define and quantify regional variations, or habitats, based on combinations of local blood flow, cell density, necrosis, and interstitial edema.

In preliminary work [7, 9], we developed image analytic methods to combine imaging sequences usually including gadolinium contrast enhanced T1 sequence (T1Gd) reflecting vascularity and breaches in blood–brain barrier, i.e., contrast enhancement, fluid-attenuated inversion recovery (FLAIR) sensitive to interstitial edema, and Apparent Diffusion Coefficient (ADC) sensitive to cell density. On MRI, glioblastoma presents as a centrally enhancing mass lesion often containing variable amounts of central

Color Code

Red: High T1PG, High Flair
Blue: Low T1PG, Low Flair
Yellow: High T1PG, Low Flair
Pink: Low T1PG, High Flair
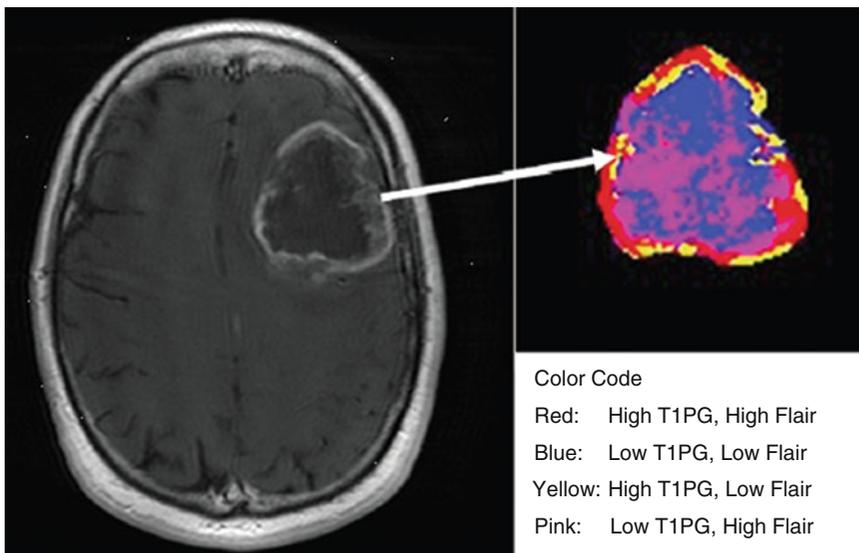
Fig. 5 *Left panel*: Contrast-enhanced T1 image from subject TCGA-02-0034 in The Cancer Genome Atlas-Glioblastoma Multiforme repository of MR volumes of GBM cases. *Right panel*: Spatial distribution of MRI-defined habitats within the tumor. The *blue region* (Low T1 post-gadolinium, low FLAIR) is particularly notable because it presumably represents a habitat with low blood flow but high cell density, indicating a population presumably adapted to hypoxic, acidic conditions. Reproduced from Gatenby et al., 2013 (Ref. 9) with permission from RSNA (Color figure online)

non-enhancement. The inner parts of the tumor bounded by and including the enhancing tissue, in quantitative imaging terms, is named the "contrast enhancing region (CE)," and outside the enhancing tumor is an area of high FLAIR without contrast enhancement, radiologically and pathologically described as an area of vasogenic edema, named the "nonenhancing region (NE)." 32 glioblastoma from The Cancer Genome Atlas (TCGA) with complete survival data were used for multiparametric imaging feature analysis. Each tumor was manually segmented as the region of interest (ROI) by a radiologist, and within this ROI the normalized intensity histograms for T1Gd, FLAIR and T2 (also fluid-sensitive) sequences were made (Fig. 6) [7].

We found that the distributions of normalized intensities for all tumors in each sequence clustered around one, two or three peaks. A bimodal distribution in T1Gd sequences was the most consistent, i.e., two distinct Gaussian populations of different signal signatures. The signal intensity of 0.26 (between the normalized range of 0–1) at the trough was denoted as the threshold between the "Low" and "High" classifications of signal intensities. Then, FLAIR values were projected onto the high and low enhancement groups (Fig. 7) [7, 41]. The total number of permutations from this resulted in four distinct multiparametric combinations which produced an automated identification of "habitats," including regions of avid enhancement, low-moderate enhancement, and necrosis.

Regions of high enhancement with low FLAIR corresponded to solid areas of enhancing tumor. Regions of low enhancement with high FLAIR indicated necrosis. The most interesting regions were those with low enhancement with a low FLAIR, i.e., high cellularity, which could be interpreted as the presence of viable cell populations that have adapted to poorly perfused conditions, e.g., hypoxia and acidosis. When this spatial analysis was separately performed in the short term and long term survivors (<400 day vs. ≥400 day survival respectively), (Fig. 8) [7] the long-term survivors' tumors tended to show avid enhancement and relatively high cell density, whereas the short term survivors had a combination of necrosis and solid regions of low enhancement but a relatively high cell density. These "low enhancement, high cellularity" areas are likely to represent a treatment-resistant cell population. Based on the fraction of the tumor volume with "low enhancement and low FLAIR" signal habitats, the leave-one-out cross-validation scheme and the receiver operator curve (ROC) demonstrated that our imaging analysis method could separate the short term survivors from the long term survivors with 81.25 % accuracy and area under curve value (AUC) of 0.86.

These habitats were spatially mapped (Fig. 5) [9] using a nonparametric Otsu segmentation approach [42]. The Otsu algorithm iteratively searches for an optimal decision boundary until convergency, thus clustering habitats into spatial groups after an intersection operation between two MRI modalities.
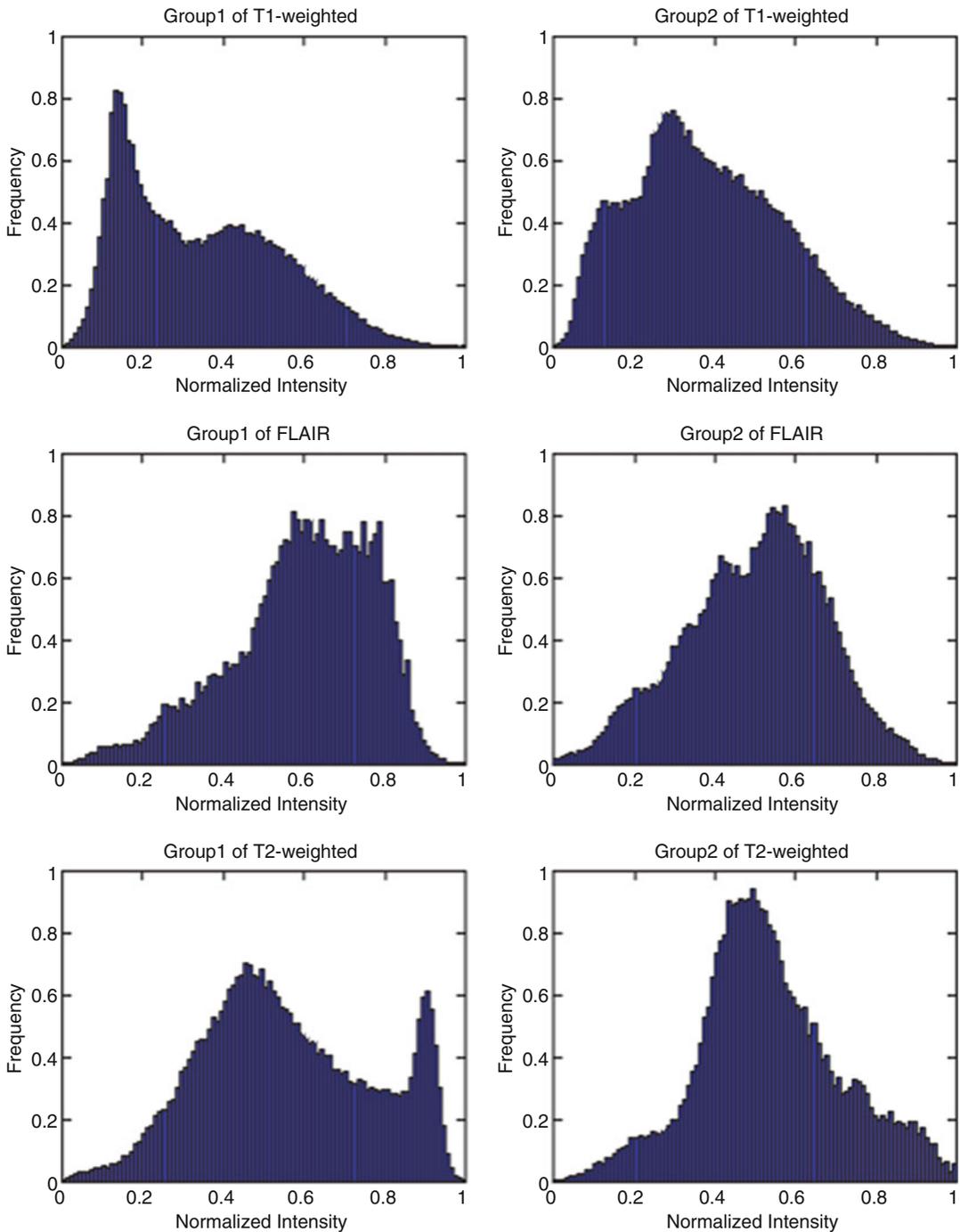
Fig. 6 Normalized cumulative two-dimensional histograms for magnetic resonance imaging intensities of TCGA glioblastoma. The normalized values (intensities) of T1 post-gadolinium, T2 weighted, and FLAIR signals were plotted for the short-term survivors (<400 days' survival, "Group 1") and long term survivors (≥400 days' survival, "Group2") respectively. Reproduced from Zhou et al., 2014 (Ref. 7) with permission from Elsevier Translational Oncology Open Access
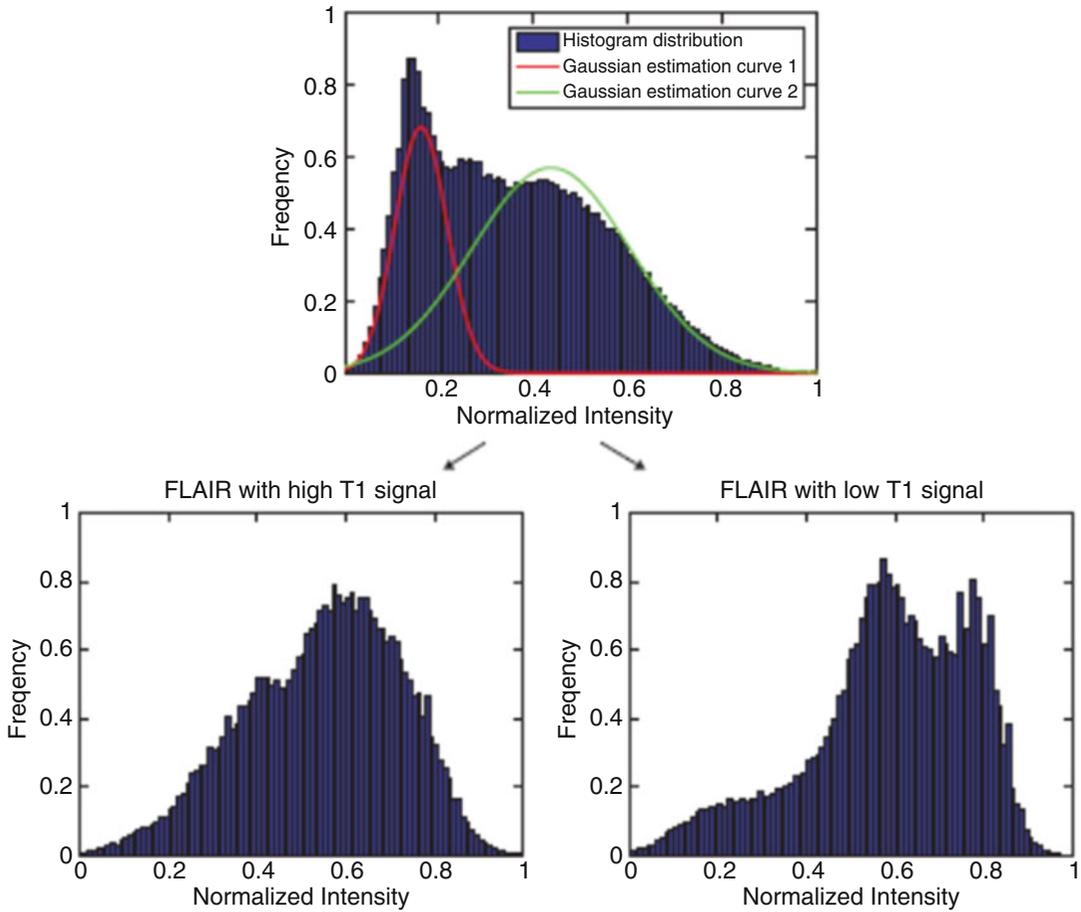
Fig. 7 *Top panel*: The frequency of the normalized values of all T1 post-gadolinium images was plotted. Using a Gaussian mixture model, the histogram was divided into two Gaussian populations with a separation point of 0.26. The normalized FLAIR signal was then plotted in the high and low T1Gd groups. *Bottom panels*: The result suggests that glioblastoma consist of five dominant habitats—two with high blood flow and high/low cell densities, and three with low blood flow and high/intermediate/low cell densities. Reproduced from Zhou et al., 2014 (Ref. 7) with permission from Elsevier Translational Oncology Open Access

8   Molecular and Histopathological Correlates of Spatially Explicit Habitats

The true merit of habitat imaging rests in its potential to facilitate individualized medicine by demonstrating targetable molecular, genetic and epigenetic characteristics of tumor cells in a noninvasive manner. At present, most efforts at expression profiling in glioma, including a large-scale study by TCGA [43], average the contributions from all local populations, and do not address the diverse cellular composition. We propose that a finer-grained characterization of regional molecular variations can lead to a better understanding of intratumoral evolutionary dynamics. In landscape ecology, the habitats noted from the satellite images are
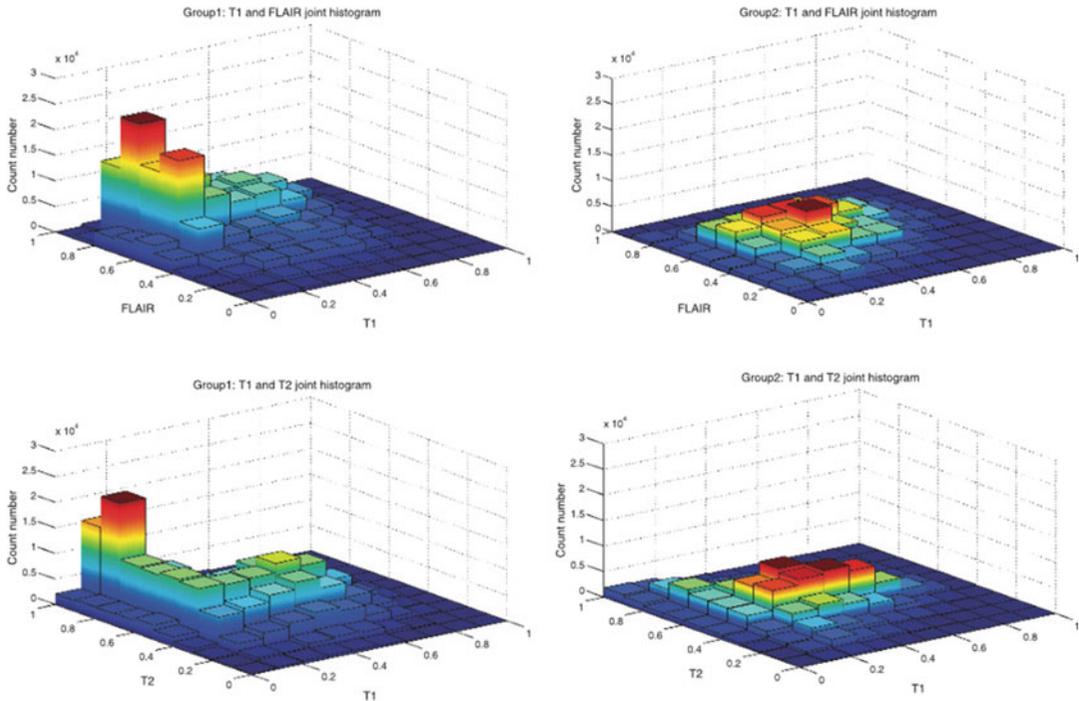
Fig. 8 Three-dimensional histograms showing the relative distribution of combinations of perfusion and cell density between short term survivors (<400 days, "Group 1") and long term survivors (≥400 days, "Group 2") of glioblastoma. For each group, we plotted the joint cumulative 3D histogram by summing all 3D histograms of each group. The long term survivors show a relatively homogeneous distribution with most regions clustering in habitats of high perfusion and intermediate cell density. The short term survivors show a greater heterogeneity with more areas of low perfusion and a mixed cell density. Reproduced from Zhou et al., 2014 (Ref. 7) with permission from Elsevier Translational Oncology Open Access

"ground-truthed" by detailed small-scale sampling and mapping of quadrats, providing a means to cross-check the heterogeneity inferred from the satellite images. In a similar fashion, in a recent Proceedings of the National Academy of Sciences (PNAS) article [17] we analyzed 69 glioblastoma patients' MRI-localized biopsy samples from the T1Gd contrast enhancing (CE) regions and the outer high FLAIR, non-enhancing infiltrative regions (NE) to produce "radio-pathomic maps" of the tumor and peritumoral regions. The pathology samples showed significant differences in the cellular density and cellular composition between the CE and NE samples. The CE had significantly higher cellularity than NE, and the CE samples were significantly more likely to contain the histological hallmarks of GBM including glomeruloid-type vascular proliferation and necrosis ($P < 0.00001$ for each feature).

The CE region also was associated with the activation of hypoxia gene expression programs. Conversely, the NE regions showed histological features of diffusely infiltrating glioma

with neoplastic glial cells intermingled with non-neoplastic and reactive cells, displaying activation of proliferation gene expression programs.

A high FLAIR signal can be interpreted as the interstitial edema, and thus, a low FLAIR signal can be assumed to indicate increased cellularity. As described above, the NE showed heterogeneity in cellularity, and a low ADC within NE correlated with a relatively high histological cell density. Interestingly, no such correlation between ADC and cell density was evident in CE tumor regions, suggesting that the addition of the ADC data to the dichotomous maps using T1Gd and FLAIR will increase the accuracy of radio-pathomic predictive models.

Consistent with ecological models, the pathological slides demonstrated a complex environment of several cell types, including glioma cells, astrocytes, neurons, microglia, and oligodendrocytes.

9   Future Directions in Quantitative Intratumoral Imaging

Can MRI signals serve as surrogate markers of the underlying cellular, molecular, tissue properties and potential drug targets? Does spatial variation in MRI features reliably correspond with pathological characteristics and prognosis? Can we identify pretreatment tumor regions and habitats that would eventually contribute disproportionately to areas of recurrence or treatment resistance? How do the radio-pathomic maps change after therapy, and do the intratumoral evolutionary patterns predict clinical outcome?

Currently we are a multidisciplinary research team incorporating expertise in neuro-radiology, neuropathology, neurosurgery, neuro-oncology, applied mathematics, statistics, systems biology, and evolutionary biology to investigate these hypotheses in over 100 glioblastoma patients enrolled in an ongoing clinical trial at Columbia University, in which each subject undergoes multiple MRI-guided biopsies prior to treatment and/or at the time of recurrence. Radiopathomic maps are constructed for each tumor in order to facilitate statistical analyses. Also, a unique set of non-surgically treated GBM patients are being investigated for the imaging features of intratumoral evolution prior to, and following therapy with temozolomide and radiotherapy, to find consistent radiologic, pathologic, or combined radio-pathomic features that predictably confer positive vs. negative clinical outcomes.

Pseudoprogression—therapy damaged brain tissue—can mimic the MRI appearance of recurrent tumor, causing significant clinical quandary during the follow up period. As our patient set includes recurrent GBMs, we examine the radio-pathomic characteristics which may provide a noninvasive answer to this difficult diagnostic problem.

10   Conclusion

Cross-sectional imaging is an indispensable part of oncological diagnosis and post-treatment surveillance; however its current utilization remains at the macroscopic level, largely limited to subjective observations. We maintain that medical imaging contains quantifiable information representing the molecular characteristics of the tissue it depicts. There is a pressing need for a superior means for objective and quantitative imaging analysis, giving way to a noninvasive imaging-based means to assess the intratumoral cellular and molecular features. The use of radiomics for imaging data extraction has provided a glimpse into the tremendous volume of information remaining to be mined and exploited in medical images. Pathologic studies have proven the presence of intratumoral molecular, genetic, and phenotypic heterogeneity, and radiomics may be the radiologic answer to tissue pathology for its ability to faithfully demonstrate spatially explicit regional variations at the microscopic level. The ecological diversity within each individual tumor is unique even amongst those belonging under the same pathologic classification, and the wide range of responses to the standard-of-care treatment is likely to be its manifestation. This phenomenon can be explained by Darwinian dynamics—each tumor as an ecosystem of distinct habitats occupied by tumor cells of a unique niche, each interacting with each other, and with the environmental selection forces. The advantages of cross-sectional imaging over tissue diagnosis are the noninvasive nature, the repeatability and its ability to interrogate the entire volume of in vivo pathology. We have presented promising evidence for pathologic proteogenomic and phenotypic features correlating to particular radiophenotypic appearances and such signatures of regions which are likely to be treatment resistant due to hypoxia are of particular interest. Quantitative imaging may enable the prediction of molecular targets for therapy and regions at risk of treatment failure before the initiation of the first round of treatment. Quantitative imaging of intratumoral evolution shows promise for contributing toward the overarching goal of individualized therapies in cancer.

References

1. Barajas RF Jr, Phillips JJ, Parvataneni R et al (2012) Regional variation in histopathologic features of tumor specimens from treatment-naive glioblastoma correlates with anatomic and physiologic MR imaging. Neuro Oncol 14:942–954

2. Gutman DA, Cooper LA, Hwang SN et al (2013) MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 267:560–569

3. Lev MH, Ozsunar Y, Henson JW et al (2004) Glial tumor grading and outcome prediction using dynamic spin-echo MR susceptibility mapping compared with conventional contrast-enhanced MR: confounding effect of elevated rCBV of oligodendrogliomas [corrected]. AJNR Am J Neuroradiol 25:214–221

4. Phillips HS, Kharbanda S, Chen R et al (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease

progression, and resemble stages in neurogenesis. Cancer Cell 9:157–173

5. Verhaak RG, Hoadley KA, Purdom E et al (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17:98–110

6. Baker M, Brown J (2010) Islands of fear: effects of wooded patches on habitat suitability of the striped mouse in a South African grassland. Funct Ecol 24:1313–1322

7. Zhou M, Hall L, Goldgof D et al (2014) Radiologically defined ecological dynamics and clinical outcomes in glioblastoma multiforme: preliminary results. Transl Oncol 7:5–13

8. Inda MM, Bonavia R, Mukasa A et al (2010) Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. Genes Dev 24:1731–1745

9. Gatenby RA, Grove O, Gillies RJ (2013) Quantitative imaging in cancer evolution and ecology. Radiology 269:8–15

10. Lacroix M, Abi-Said D, Fourney DR et al (2001) A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. J Neurosurg 95:190–198

11. Diehn M, Nardini C, Wang DS et al (2008) Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A 105:5213–5218

12. Garzon B, Emblem KE, Mouridsen K et al (2011) Multiparametric analysis of magnetic resonance images for glioma grading and patient survival time prediction. Acta Radiol 52:1052–1060

13. Pope WB, Sayre J, Perlina A et al (2005) MR imaging correlates of survival in patients with high-grade gliomas. Am J Neuroradiol 26:2466–2474

14. Aum DJ, Kim DH, Beaumont TL et al (2014) Molecular and cellular heterogeneity: the hallmark of glioblastoma. Neurosurg Focus 37:E11

15. Gerlinger M, Rowan AJ, Horswell S et al (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366:883–892

16. Gerlinger M, Swanton C (2010) How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. Br J Cancer 103:1139–1143

17. Gill BJ, Pisapia DJ, Malone HR et al (2014) MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. Proc Natl Acad Sci U S A 111:12550–12555

18. Sottoriva A, Spiteri I, Piccirillo SG et al (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A 110:4009–4014

19. Yancovitz M, Litterman A, Yoon J et al (2012) Intra- and inter-tumor heterogeneity of BRAF(V600E) mutations in primary and metastatic melanoma. PLoS One 7:e29336

20. Patel AP, Tirosh I, Trombetta JJ et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344:1396–1401

21. Zar JH (2009) Biostatistical analysis, 5th edn. Pearsons

22. Bisdas S, Kirkpatrick M, Giglio P et al (2009) Cerebral blood volume measurements by perfusion-weighted MR imaging in gliomas: ready for prime time in predicting short-term outcome and recurrent disease? Am J Neuroradiol 30:681–688

23. Jain R, Poisson L, Narang J et al (2013) Genomic mapping and survival prediction in glioblastoma: molecular subclassification strengthened by hemodynamic imaging biomarkers. Radiology 267:212–220

24. Jain R, Poisson L, Narang J et al (2012) Correlation of perfusion parameters with genes related to angiogenesis regulation in glioblastoma: a feasibility study. Am J Neuroradiol 33:1343–1348

25. Law M, Young RJ, Babb JS et al (2008) Gliomas: predicting time to progression or survival with cerebral blood volume measurements at dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. Radiology 247:490–498

26. Mills SJ, Patankar TA, Haroon HA et al (2006) Do cerebral blood volume and contrast transfer coefficient predict prognosis in human glioma? Am J Neuroradiol 27:853–858

27. Pope WB, Chen JH, Dong J et al (2008) Relationship between gene expression and enhancement in glioblastoma multiforme: exploratory DNA microarray analysis. Radiology 249:268–277

28. Swanson KR, Rockne RC, Claridge J et al (2011) Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology. Cancer Res 71:7366–7375

29. Van Meter T, Dumur C, Hafez N et al (2006) Microarray analysis of MRI-defined tissue samples in glioblastoma reveals differences in regional expression of therapeutic targets. Diagn Mol Pathol 15:195–205

30. Balagurunathan Y, Gu Y, Wang H et al (2014) Reproducibility and prognosis of quantitative

features extracted from CT images. Transl Oncol 7:72–87

31. Balagurunathan Y, Kumar V, Gu Y et al (2014) Test-retest reproducibility analysis of lung CT image features. J Digit Imaging 27:805–823

32. Colen R, Foster I, Gatenby R et al (2014) NCI workshop report: clinical and computational requirements for correlating imaging phenotypes with genomics signatures. Transl Oncol 7:556–569

33. Gu Y, Kumar V, Hall LO et al (2013) Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. Pattern Recognit 46:692–702

34. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248

35. Jain RK (2005) Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science 307:58–62

36. Wang Z, Dabrosin C, Yin X et al (2015) Broad targeting of angiogenesis for cancer prevention and therapy. Semin Cancer Biol 35(Suppl): S224–S243

37. Ehdaie B, Vertosick E, Spaliviero M et al (2014) The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. J Urol 191:660–664

38. Kern SE (2012) Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res 72:6097–6101

39. Mirnezami R, Nicholson J, Darzi A (2012) Preparing for precision medicine. N Engl J Med 366:489–491

40. Gatenby RA, Gillies RJ (2008) A microenvironmental model of carcinogenesis. Nat Rev Cancer 8:56–61

41. Sonoda Y, Kumabe T, Watanabe M et al (2009) Long-term survivors of glioblastoma: clinical features and molecular analysis. Acta Neurochir 151:1349–1358

42. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9:62–66

43. Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068

Chapter 7

Transcriptome and Proteome Analyses of TNFAIP8 Knockdown Cancer Cells Reveal New Insights into Molecular Determinants of Cell Survival and Tumor Progression

Timothy F. Day, Rajshree R. Mewani, Joshua Starr, Xin Li, Debyani Chakravarty, Habtom Ressom, Xiaojun Zou, Ofer Eidelman, Harvey B. Pollard, and Meera Srivastava, and Usha N. Kasid

Abstract

Tumor necrosis factor-α-inducible protein 8 (TNFAIP8) is the first discovered oncogenic and an anti-apoptotic member of a conserved TNFAIP8 or TIPE family of proteins. TNFAIP8 mRNA is induced by NF-kB, and overexpression of TNFAIP8 has been correlated with poor prognosis in many cancers. Downregulation of TNFAIP8 expression has been associated with decreased pulmonary colonization of human tumor cells, and enhanced sensitivities of tumor xenografts to radiation and docetaxel. Here we have investigated the effects of depletion of TNFAIP8 on the mRNA, microRNA and protein expression profiles in prostate and breast cancers and melanoma. Depending on the tumor cell type, knockdown of TNFAIP8 was found to be associated with increased mRNA expression of several antiproliferative and apoptotic genes (e.g., IL-24, FAT3, LPHN2, EPHA3) and fatty acid oxidation gene ACADL, and decreased mRNA levels of oncogenes (e.g., NFAT5, MALAT1, MET, FOXA1, KRAS, S100P, OSTF1) and glutamate transporter gene SLC1A1. TNFAIP8 knockdown cells also exhibited decreased expression of multiple onco-proteins (e.g., PIK3CA, SRC, EGFR, IL5, ABL1, GAP43), and increased expression of the orphan nuclear receptor NR4A1 and alpha 1 adaptin subunit of the adaptor-related protein complex 2 AP2 critical to clathrin-mediated endocytosis. TNFAIP8-centric molecules were found to be predominately implicated in the hypoxia-inducible factor-1α (HIF-1α) signaling pathway, and cancer and development signaling networks. Thus TNFAIP8 seems to regulate the cell survival and cancer progression processes in a multifaceted manner. Future validation of the molecules identified in this study is likely to lead to new subset of molecules and functional determinants of cancer cell survival and progression.

Key words TNFAIP8, shRNA and siRNA, RNA and microRNA arrays, Antibody arrays, Cell survival and proliferation, Invasion and metastasis, Cancer systems biology

1   Introduction

The tumor necrosis factor-α-induced protein 8 (TNFAIP8 or TIPE) family of proteins is comprised of four members, TNFAIP8, TNFAIP8-like 1 (TIPE1), TNFAIP8-like 2 (TIPE2),

and TNFAIP8-like 3 (TIPE3) [1–7]. These proteins have a conserved TIPE homology (TH) domain containing a large hydrophobic cavity [8]. TIPE members bind to phosphoinositide 4,5-bisphosphate (PtdIns(4,5)P_2) and phosphoinositide 3,4,5-bisphosphate (PtdIns(3,4,5)P_3) and may represent a new class of phospholipid transfer proteins [6]. Specifically, TIPE3 has been shown to increase the levels of PtdIns(4,5)P_2 and PtdIns(3–5)P_3 and constitutes a novel phosphoinositide metabolism regulator [6]. The members of the TIPE family function as pro-apoptotic molecules (TIPE1 and TIPE 2) or pro-survival and oncogenic molecules (TNFAIP8 and TIPE3) [2–14].

1.1 TNFAIP8 Is a Novel Pro-survival, Oncogenic, and Metastatic Molecule

TNFAIP8 is the first discovered member of the TIPE family of proteins [1, 2, 9, 10]. It was identified as a novel transcript overexpressed in metastatic and radioresistant head and neck squamous cell carcinoma cells as compared to their matched primary tumor-derived cells [1]. Expression analysis showed high TNFAIP8 mRNA expression in thyroid, lymph nodes, and bone marrow, and very low expression in other normal tissues such as the lung and liver [2]. TNFAIP8 is a predominantly cytosolic molecule [11]. There are four major isoforms of TNFAIP8 (23, 22, 21.8, and 24 kDa) (www.uniprot.org) (Figs. 1a, b) [14]. TNFAIP8 isoform 1 (23 kDa) and isoform 3 (aliases SCC-S2, GG2-1, NDED, MDC3.13, 21.8 kDa) (21.8 kDa) are highly expressed in most cancers cell lines [11, 12]. TNFAIP8 mRNA and protein expression is induced by tumor necrosis factor α (TNF-α) in human tumor cells and endothelial cells [2, 9]. In addition, TNFAIP8 expression is induced by NF-kB and overexpression of TNFAIP8 promotes cell survival in NF-kB null cells, suggesting that it is an important pro-survival molecule downstream of NF-kB [10]. TNFAIP8 overexpression was also found to inhibit caspase-8 activity further supporting its pro-survival function [10].

The promoter analysis of an approximately 2.0 kb genomic sequence upstream of the translation start site of *TNFAIP8* gene indicates potential binding sites for multiple transcription factors including NF-kB, hypoxia-inducible factor (HIF), COUP-TFI, and androgen receptor [13, 15]. A complex including the pro-apoptotic protein DBC1 and COUP-TF1 was found to occupy the *TNFAIP8* promoter. This complex was found to be crucial for repression of the *TNFAIP8* promoter by COUP-TFI [15]. TNFAIP8 expression is also induced in response to androgen in prostate cancer cells [13]. In "promoter array" studies, differential methylation and regulation of TNFAIP8 has been reported in prostate epithelial and cancer cell lines [16].

TNFAIP8 is an oncogenic and metastatic molecule. Overexpression of TNFAIP8 increases cell proliferation in MCF-7 breast cancer cells (Fig. 2) [17], and cell proliferation, and migration and invasion in MDA-MB-435 melanoma cells [11, 12].

A

Variant 1 (2103 bp, includes 5'UTR 271 bp and 3'UTR 1238 bp)

Variant (X) (2003 bp, includes 5' UTR 197 bp and 3' UTR 1236 bp

Variant 2 (2017 bp, includes 5'UTR 215 bp and 3'UTR 1238 bp)

Variant 4 (2258 bp, includes 5'UTR 390 bp and 3'UTR 1230 bp)
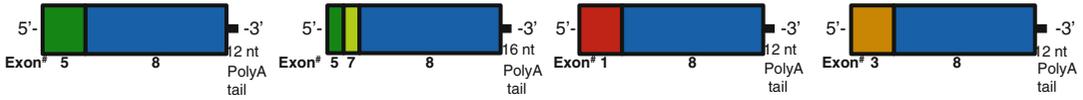
Protein Isoform 1 (198 aa, 23.0 kDa) MDC-3.13-2 (Uniprot: O95379-1)

MHSEAEESKEVATDVFNSKNL AVQAQKKILGKMVSKSIATTLI DDTSSEVLDELYRVTREYTQN KKEAEKIIKNLIKTVIKLAILYRN NQFNQDELALMEKFKKKVHQ LAMTVVSFHQVDYTFDRNVLS RLLNECREMLHQIIQRHLTAKS HGRVNNVFDHFSDCEFLAAL YNPFGNFKPHLQKLCDGINKM LDEENI

Protein Isoform 2 (190 aa, 22 kDa) MDC-3.13-1 (Uniprot: O95379-2)

MAVATDVFNSKNLAVQAQKKI LGKMVSKSIATTLIDDTSSEVL DELYRVTREYTQNKKEAEKIIK NLIKTVIKLAILYRNNQFNQDE LALMEKFKKKVHQLAMTVVSF HQVDYTFDRNVLSRLLNECR EMLHQIIQRHLTAKSHGRVNN VFDHFSDCEFLAALYNPFGNF KPHLQKLCDG INKMLDEENI

Protein Isoform 3 (188 aa, 21.8 kDa) SCC-S2, GG2-1, NDED, MDC-3.13 (Uniprot: O95379-3)

MATDVFNSKNLAVQAQKKILG KMVSKSIATTLIDDTSSEVLDE LYRVTREYTQNKKEAEKIIKNLI KTVIKLAILYRNNQFNQDELAL MEKFKKKVHQLAMTVVSFHQ VDYTFDRNVLSRLLNECREML HQIIQRHLTAKSHGRVNNVFD HFSDCEFLAALYNPFGNFKPH LQKLCDGINMLDEENI

Protein Isoform 4 (210 aa, 24.4 kDa) (Uniprot: O95379-4)

MTLPRYCEVVLLIAHGEKMLK LVATDVFNSKNLAVQAQKKIL GKMVSKSIATTLIDDTSSEVLD ELYRVTREYTQNKKEAEKIIKN LIKTVIKLAILYRNNQFNQDELA LMEKFKKKVHQLAMTVVSFH QVDYTFDRNVLSRLLNECRE MLHQIIQRHLTAKSHGRVNNV FDHFSDCEFLAALYNPFGNFK PHLQKLCDGINKMLDEENI

B

```
Isoform1    1   ------------MHSEAEESKEVATDVFNSKNLAVQAQKKILGKMVSKSIATTLIDDTSS   48
Isoform2    1   --------------------MAVATDVFNSKNLAVQAQKKILGKMVSKSIATTLIDDTSS   40
Isoform3    1   ----------------------MATDVFNSKNLAVQAQKKILGKMVSKSIATTLIDDTSS   38
Isoform4    1   MTLPRYCEVVLLIAHGEKMLKLVATDVFNSKNLAVQAQKKILGKMVSKSIATTLIDDTSS   60
                                    .:*********************************************

Isoform1   49   EVLDELYRVTREYTQNKKEAEKIIKNLIKTVIKLAILYRNNQFNQDELALMEKFKKKVHQ  108
Isoform2   41   EVLDELYRVTREYTQNKKEAEKIIKNLIKTVIKLAILYRNNQFNQDELALMEKFKKKVHQ  100
Isoform3   39   EVLDELYRVTREYTQNKKEAEKIIKNLIKTVIKLAILYRNNQFNQDELALMEKFKKKVHQ   98
Isoform4   61   EVLDELYRVTREYTQNKKEAEKIIKNLIKTVIKLAILYRNNQFNQDELALMEKFKKKVHQ  120
                ***********************************************************

Isoform1  109   LAMTVVSFHQVDYTFDRNVLSRLLNECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFL  168
Isoform2  101   LAMTVVSFHQVDYTFDRNVLSRLLNECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFL  160
Isoform3   99   LAMTVVSFHQVDYTFDRNVLSRLLNECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFL  158
Isoform4  121   LAMTVVSFHQVDYTFDRNVLSRLLNECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFL  180
                ***********************************************************

Isoform1  169   AALYNPFGNFKPHLQKLCDGINKMLDEENI  198
Isoform2  161   AALYNPFGNFKPHLQKLCDGINKMLDEENI  190
Isoform3  159   AALYNPFGNFKPHLQKLCDGINKMLDEENI  188
Isoform4  181   AALYNPFGNFKPHLQKLCDGINKMLDEENI  210
```

Fig. 1 (**a**) Human TNFAIP8: predicted alternative transcripts including coding and untranslated sequences and corresponding protein isoforms are shown. (**b**) Alignment of the amino acid sequences of various human TNFAIP8 isoforms. Protein isoform sequences were obtained from Uniprot website (http://www.uniprot.org/uniprot/O95379) as described earlier [14]
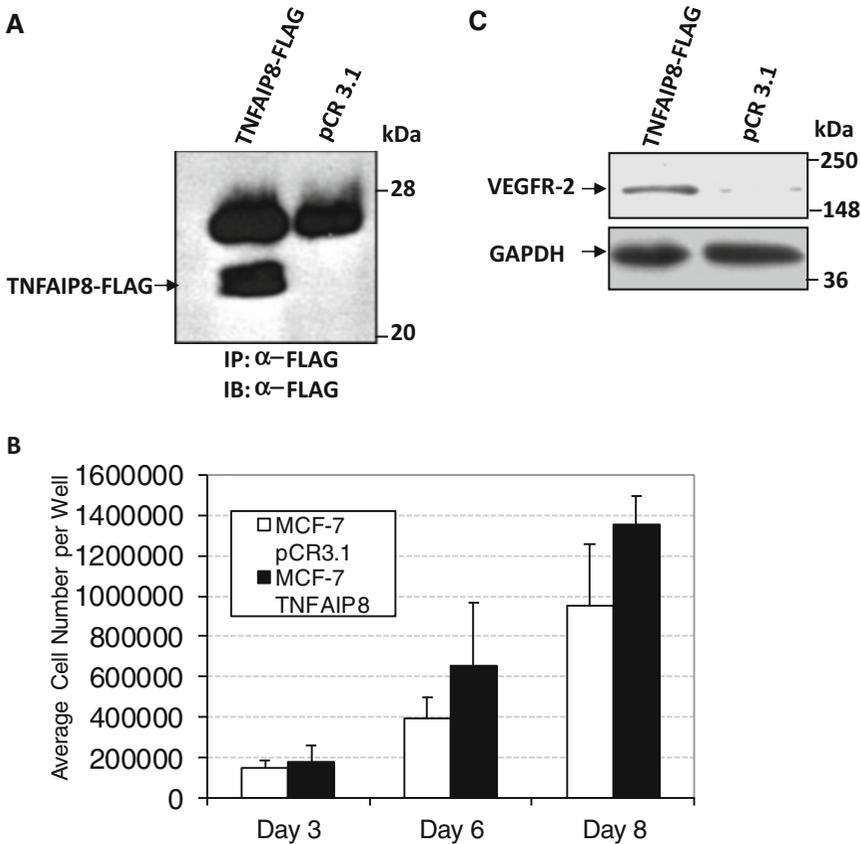
Fig. 2 Effects of TNFAIP8 overexpression on cell proliferation and VEGFR-2 expression in MCF-7 breast cancer cells. (**a**) Expression of FLAG-tagged TNFAIP8 in stably transfected MCF-7 cells was examined as described earlier [17]. *IP* immunoprecipitation, *IB* immunoblotting. (**b**) Enhanced cell proliferation in MCF-7 TNFAIP8 transfectants. TNFAIP8 and pCR3.1 vector transfectants were seeded in six well plates (5×10^4 per well in triplicate), and viable cell numbers on various days were counted by the trypan blue dye exclusion method as described earlier [17]. Values represent the average cell number ± S.E. from three independent experiments. (**c**) Enhanced VEGFR-2 protein expression in MCF-7 TNFAIP8 stable transfectants. Cells were lysed and whole cell lysates were analyzed by Western blotting using anti-VEGFR-2 antibody as explained earlier [17]. The blot was reprobed with anti-GAPDH antibody. Approximately ninefold increase in VEGFR-2 expression was seen in MCF-7 TNFAIP8 transfectants versus control vector transfectants

Consistently, siRNA downregulation of TNFAIP8 decreases invasion in MDA-MB-435 cells (Fig. 3) [17]. Furthermore, injection of TNFAIP8 overexpressing cells into mammary fat pads in athymic female mice led to accelerated tumor formation as compared to control cells [11]. In other studies, tail-vein injection of MDA-MB-435 cells stably expressing TNFAIP8 led to significant pulmonary colonization of tumor cells as compared to control cells [12]. In addition, systemic delivery of liposome-entrapped TNFAIP8 targeted antisense oligos resulted in significant decrease in pulmonary colonization of MDA-MB-435 cells in athymic mice as compared to mice receiving control liposomes [12].

A

C

Scrambled siRNA (Scr1)   TNFAIP8 siRNA (Si2)

B

D
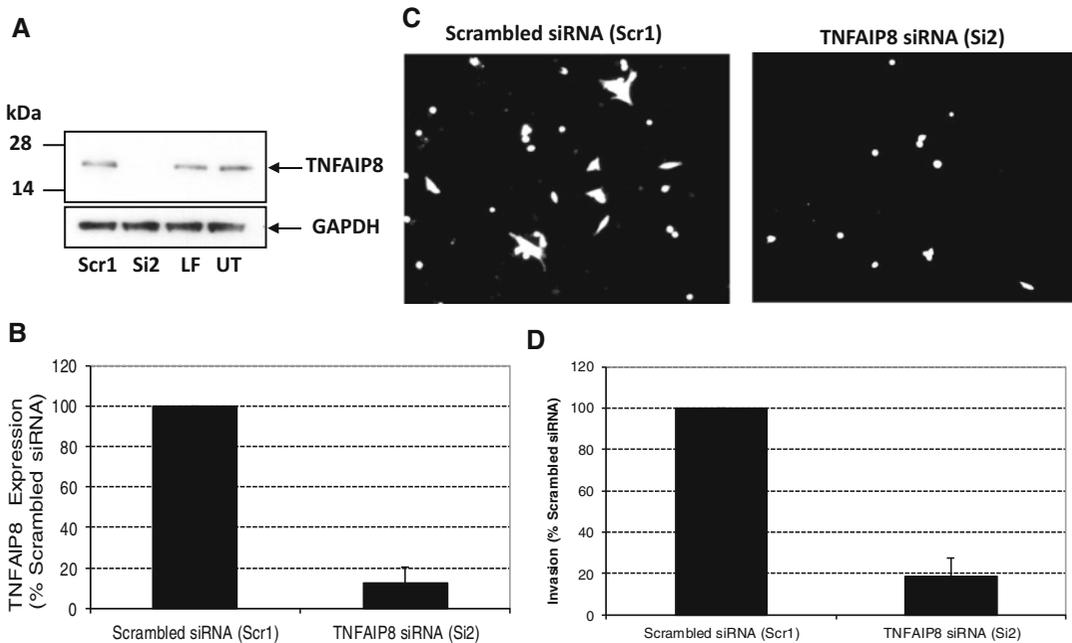
Fig. 3 SiRNA Knockdown of TNFAIP8 protein expression is associated with decreased invasiveness in MDA-MB435 melanoma cells. (**a**) MDA-MB 435 cells were treated with TNFAIP8 siRNA (Si2, 200 nM) or a scrambled siRNA (Scr1, 200 nM) and TNFAIP8 expression was determined by Western blotting using anti-TNFAIP8 antibody as described earlier [12, 13, 17]. The blot was reprobed with anti-GAPDH antibody. The TNFAIP8 signal was quantified and normalized against GAPDH in the same lane. *UT* untreated control cells, *LF* Lipofectamine 2000 treated control cells. (**b**) Quantification of normalized TNFAIP8 expression was performed using ImageQuant software. Data shown are mean ± S.D. Approximately 88 % of TNFAIP8 expression was inhibited in the siRNA treated cells (Si2) versus Scr control cells (Scr1) ($N=3$, $p < 0.00003$) as described earlier [17]. (**c**) MDA-MB435 cells were treated with TNFAIP8 siRNA (Si2, 200 nM) or scrambled siRNA (Scr1, 200 nM) and modified Boyden chamber invasion assay was performed and data were quantified as described earlier [17, 39]. Representative images of the invaded cells as seen under the TE300X inverted microscope are shown. (**d**) TNFAIP8 siRNA treated MDA-MB435 cells show decrease in the number of invasive cells. Data shown is mean ± S.D. Approximately 81 % of the siRNA treated cells (Si2) showed loss of invasiveness versus Scr control cells (Scr1) ($N=3$, $p < 0.00005$) as described earlier [17]. The TNFAIP8 Stealth™ siRNA duplex oligoribonucleotide sequence used in the present study was Si2, 5′- UAAAUUUCUCCAUCAAUGCUAGCUC-3′ (TNFAIP8-HSS119330, Invitrogen). The control siRNAs used was Scrambled #1 (Stealth RNAi Negative Control Low GC Duplex; Invitrogen; 12935-200)

1.2 TNFAIP8 Is Implicated in Cancer Prognosis and Therapy Response

TNFAIP8 appears to be an important cancer biomarker. Numerous publicly available datasets (ONCOMINE, GEO) show correlation between high TNFAIP8 expression and hyperplasia, primary tumors, higher grade, and/or metastatic lesions of diverse organs including prostate, kidney, liver, breast, head-neck, brain, lung, melanoma, and pancreas [13]. Furthermore, immunohistochemical studies have reported significantly higher TNFAIP8 protein expression in multiple tumor types. These include NSCLC, pancreatic carcinoma, prostate carcinoma, and thyroid cancer [13, 18–20].

Expression of TNFAIP8 has been associated with increased resistance to chemotherapeutic drugs cytosine arabinoside and anthracyclines in acute myeloid leukemia [21]. Furthermore, anti-sense downregulation of TNFAIP8 has been associated with enhanced sensitivities to paclitaxel and docetaxel in prostate tumor xenografts grown in athymic mice [13]. Moreover, nuclear TNFAIP8 expression has been correlated with worse prognosis in patients with prostatic adenocarcinomas [13]. Taken together, these observations suggest that TNFAIP8 is a viable therapeutic target in cancer.

1.3 Interacting Partners of TNFAIP8

Mechanism of TNFAIP8 action remains unclear. In melanoma and prostate cancer cells, antisense knockdown of TNFAIP8 correlated with concomitant decrease in expression of angiogenic and meta-static molecules such as VEGFR2, matrix metalloproteinase 1 (MMP1), and MMP9 [12]. Depletion of TNFAIP8 also corre-lated with decreased expression of VEGFR2 in endothelial cells [12] and MCF-7 breast cancer cells (Fig. 2b) [17]. These data suggest that TNFAIP8 may be a regulator of the VEGF-signaling and associated biological response. In pancreatic adenocarcinoma, TNFAIP8 expression was found to strongly correlate with increased EGFR expression [19]. In yeast two hybrid studies, TNFAIP8 was found to interact with Galpha(i) and inhibit cell death in caspase-independent manner in Balb-D2S cells [22]. Other proteins inter-acting with TNFAIP8 include GDNF family receptor α 1, PARP1, Serine/arginine-rich splicing factor 2, Karyopherin α2 (RAG cohort 1, importin α1), DEAD (Asp-Glu-Ala-Asp) box polypep-tide 20, protein tyrosine phosphatase, non-receptor type 1, and cyclin E1 [13]. Further investigations are necessary to define a functional link between TNFAIP8 and its interacting partners.

In this chapter, we have investigated the effects of shRNA or siRNA knockdown of TNFAIP8 expression on genome-wide changes in the mRNA, microRNA (miR), and protein expression profiles in cancer cells. We have identified a subset of molecules that may reflect the biology of TNFAIP8 in various cancer cell types. Ultimately this work is likely to advance knowledge of the signaling molecules and networks that could determine cell sur-vival and proliferation, and tumor progression behavior in certain cancers.

2 Effects of TNFAIP8 Knockdown on the mRNA and microRNA Expression Profiles in Cancer Cells

Several human prostate, breast, pancreatic, and lung cancer cell lines were stably transfected with TNFAIP8 shRNA or scrambled control vector using the pLKO.1 lentiviral transduction procedure as described earlier [14]. The knockdown of TNFAIP8 expression
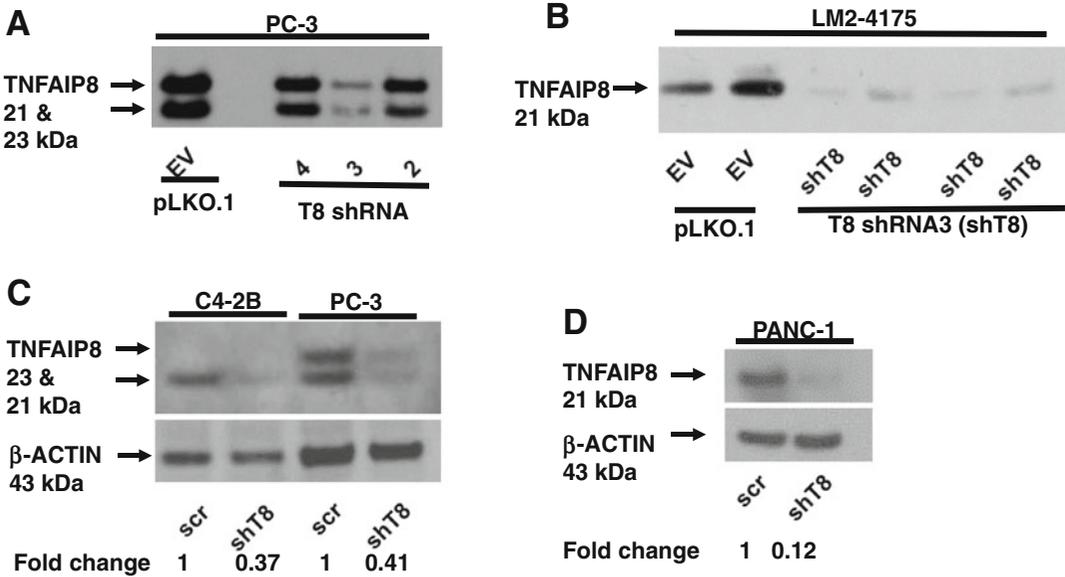
Fig. 4 Silencing of TNFAIP8 expression by stable transduction of lentiviral TNFAIP8 shRNA expression vector into various cancer cell lines as described earlier [14]. (**a**) and (**b**) Immunoblot analysis of TNFAIP8 knockdown in PC-3 prostate cancer cells and LM2-4175 breast cancer cells stably transfected with indicated TNFAIP8 shRNA (T8 shRNA #s 2, 3 and 4). Controls cells were stably transfected with empty vector (EV). (**c**) and (**d**) Immunoblot analysis of TNFAIP8 knockdown in prostate (C4-2B and PC-3) and pancreatic cancer cell lines (PANC-1) stably transfected with lentiviral TNFAIP8 shRNA # 3 expression vector (shT8). Scr, scrambled shRNA (scr). The blots were reprobed with anti-β-Actin antibody and the fold change in TNFAIP8 expression relative to scr control was obtained after normalizing against β-ACTIN expression in the corresponding lanes

in various tumor cell models was verified by the RT-PCR, qRT-PCR, and Western blotting assays. Representative data are shown in Fig. 4. For mRNA profiling, total RNA was isolated from the PC-3 and C4-2B prostate cancer models of TNFAIP8 knockdown (shT8) and control cells (scr) cells. Samples were processed in triplicate using RNA arrays (Affymetrix Human Genome U133 Plus 2.0 arrays). Data were processed and quantified, followed by the Ingenuity Pathway Analysis (IPA) of the top-ranked expression patterns (≥2-fold up/down) in TNFAIP8 shRNA versus scrambled shRNA treated cells. Nineteen distinct mRNAs that were modulated in a TNFAIP8 knockdown-specific manner in prostate cancer cells are shown in Tables 1 and 2. Fifteen of these mRNAs were validated by qRT-PCR in multiple tumor model cell systems (Fig. 5). In brief, TNFAIP8 knockdown tumor cells showed increased mRNA expression of tumor suppressor genes such as IL24 (mda-7), FAT3, LPHN2, EPHA3 and IGFBP3 (Fig. 5a), and decreased expression of tumor progression markers, e.g., NFAT5, MALAT1, OSTF1, FOXA1, KRAS, MAP2K6, S100P, FLRT2, and MET (Fig. 5b). Interestingly, expression of ACADL, long chain acyl-CoA dehydrogenase was significantly increased

Table 1
Selected RNA array data showing changes in mRNA expression in TNFAIP8 knockdown PC-3 prostate cancer cells

Gene symbol	Description[a]	Fold change in mRNA in shTNFAIP8 vs Scr PC-3 cells (+, high in shTNFAIP8; −, low in shTNFAIP8)[b]
IL24 (mda-7)	Interleukin 24	+6.3
FAT3	FAT tumor suppressor homolog 3 (Drosophila)	+6.3
LPHN2	Latrophilin 2	+5.9
EPHA3	EPH receptor A3	+3.4
TMED4	**Transmembrane emp24 protein transport domain containing 4**	**+2.3**
DHRS2	**Dehydrogenase/reductase (SDR family) member 2**	**+2.0**
SNX1	Sorting nexin 1	+1.4
NFAT5	Nuclear factor of activated T-cells 5, tonicity-responsive	−9.1
MALAT1	Metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	−5.9
MET	Met proto-oncogene (hepatocyte growth factor receptor)	−3.4
FOXA1	Forkhead box A1	−3.1
TPR	**Translocated promoter region (to activated MET oncogene)**	**−3.1**
TNFAIP8	Tumor necrosis factor, alpha-induced protein 8	−2.9
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	−2.2
SLC1A1	**Solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1**	**−2.2**
MAP2K6	Mitogen-activated protein kinase kinase 6	−2.1

[a]*Bold* indicates that similar changes in mRNA expression were observed in TNFAIP8 knockdown C4-2B cells
[b]*P*-value, <0.05

(Table 2 and Fig. 5a), and expression of SLC1A1, solute carrier family 1 high affinity glutamate transporter gene, was decreased in TNFAIP8 knockdown cells (Tables 1 and 2).

For the microRNA expression profiling, RNA was isolated from the PC-3 and C4-2B prostate cancer models of TNFAIP8 knockdown (shT8) and control cells (scr) cells using the RNA isolation kit (Qiagen). Samples were processed in triplicate using microRNA arrays (MIRNA 2.0, Affymetrix Inc.). Data were processed and quantified. Changes in human microRNA levels in

Table 2
Selected RNA array data showing changes in mRNA expression in TNFAIP8 knockdown C4-2B prostate cancer cells

Gene symbol	Description[a]	Fold change in mRNA in shTNFAIP8 vs Scr C4-2B cells (+, high in shTNFAIP8; −, low in shTNFAIP8)[b](FDR)
ACADL	Acyl-CoA dehydrogenase, long chain	+15.1 (0.00729)
IGFBP3	Insulin-like growth factor binding protein 3	+5.2 (0.0432)
TMED4	**Transmembrane emp24 protein transport domain containing 4**	**+2.6 (0.00729)**
DHRS2	**Dehydrogenase/reductase (SDR family) member 2**	**+2.0 (0.0365)**
S100P	S100 calcium binding protein P	−7.7 (0.0219)
FLRT2	Fibronectin leucine rich transmembrane protein 2	−4.0 (0.0859)
TNFAIP8	Tumor necrosis factor, alpha-induced protein 8	−2.6 (0.0395)
SLC1A1	**Solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1**	**−2.6 (0.0458)**
OSTF1	Osteoclast stimulating factor 1	−2.4 (<1e-07)
TPR	**Translocated promoter region (to activated MET oncogene)**	**−2.3**

[a]*Bold* indicates that similar changes in mRNA expression were observed in TNFAIP8 knockdown PC-3 cells
[b]*P*-value, < 0.05; FDR (False Discovery Rate, < 0.1 values are shown)

TNFAIP8 depleted cells relative to control cells are shown in Table 3. Verification of various microRNAs in multiple tumor cell models of increased and decreased expression of TNFAIP8, and functional validation of the top microRNAs and their targets are currently ongoing in our laboratory.

On the PC-3 and C4-2B mRNA array datasets, "core" analyses were run with a fold change cutoff of ±1.3. Fisher's *P*-values were given as a measure of overlap between observed and predicted gene sets. Top connectivity maps, top canonical pathways and top diseases and biofunctions were generated by Ingenuity Pathway Analysis (IPA). Based on the integrated transcriptome analysis, aryl hydrocarbon receptor signaling and ephrin receptor signaling are among the top-ranked canonical pathways, and cancer and tumor morphology are among the top-ranked diseases and biofunctions signifying TNFAIP8-centric changes in transcriptome of cancer cells (Fig. 6, and Tables 4 and 5).

A
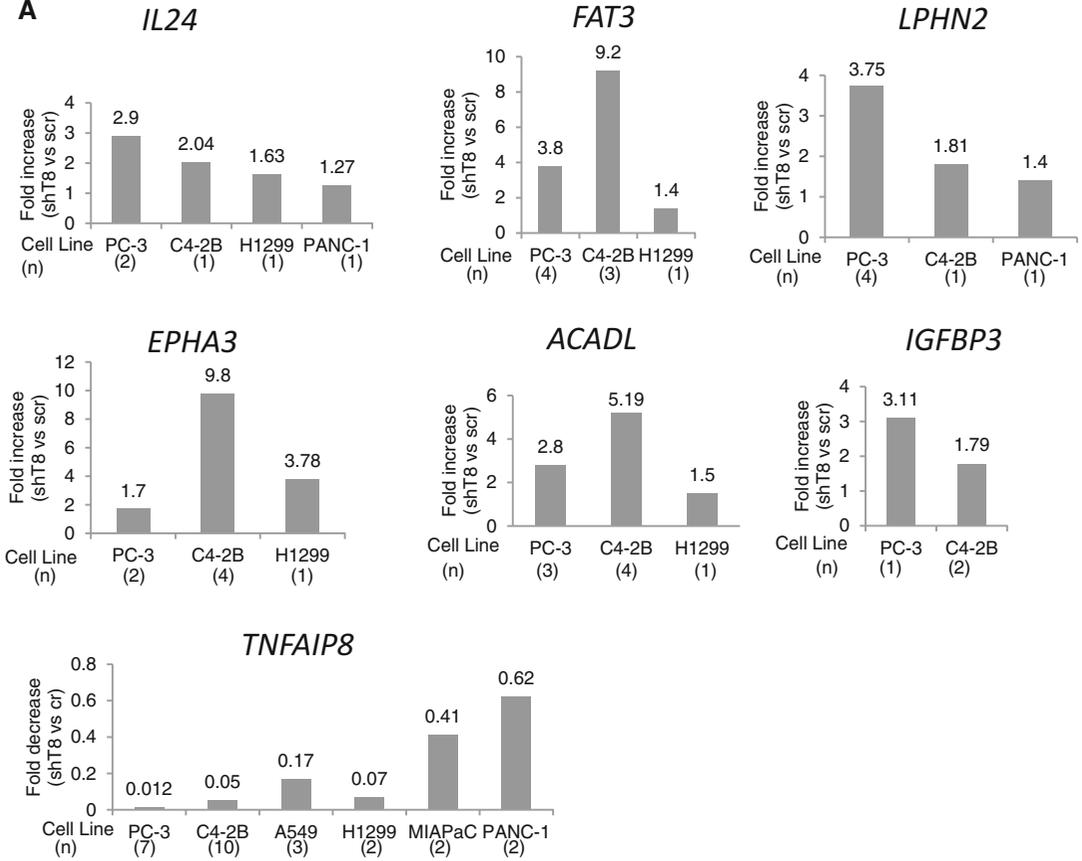
B
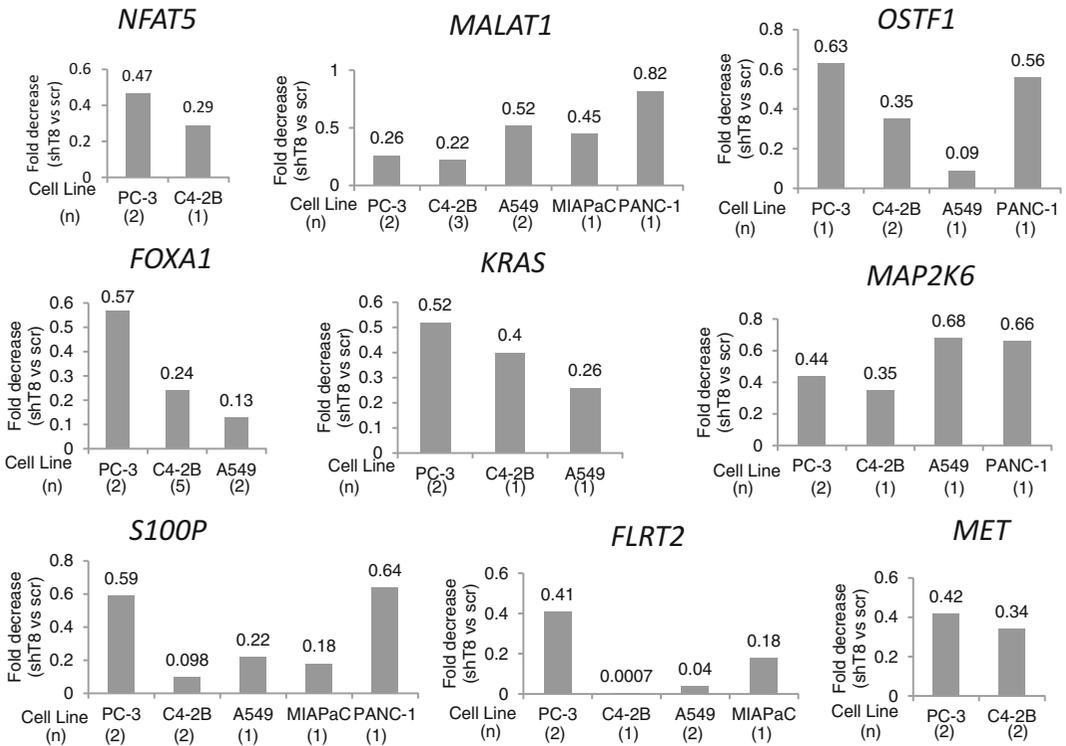
Table 3
Selected microRNA array data showing changes in human miR expression in shTNFAIP8 knockdown PC-3 and C4-2B prostate cancer cells

Human microRNA	Fold change in mRNA in shTNFAIP8 vs Scr cells (+, high in shTNFAIP8; −, low in shTNFAIP8)	*P*-value
I. PC-3		
hsa-miR-3172	+4.0	0.008
hsa-miR-1274a	+3.0	0.003
hsa-mir-429	+2.4	0.036
hsa-miR-3180-3p	+2.3	0.013
hsa-miR-720	+2.1	0.049
hsa-miR-1274b	+2.1	0.004
hsa-miR-1975	+2.0	0.024
hsa-miR-767	−2.7	0.013
hsa-miR-514b	−2.5	0.000
hsa-mir-221	−2.2	0.001
hsa-miR-943	−2.1	0.000
II. C4-2B		
hsa-miR-7	+3.8	0.008
hsa-miR-29b-1-star	+2.2	0.021

3 Effects of TNFAIP8 Depletion on the Protein Expression Profiles in Cancer Cells

The protein expression profiles were investigated in TNFAIP8 knockdown models of prostate (PC-3), breast (MDA-MB-231 and LM2-4175) and melanoma cell lines (MDA-MB435) using antibody microarrays (Antibody Microarray 507, Clontech)

Fig. 5 Validation of changes in mRNA expression in TNFAIP8 knockdown cells by qRT-PCR. (**a**) Increased mRNA expression of indicated genes in shTNFAIP8 (shT8) knockdown models of human prostate (PC-3, C4-2B), lung (A549, H1299), and pancreatic cancer cell lines (MIAPaC, PANC-1) versus scrambled control cells (scr). (**b**) Decreased mRNA expression of indicated genes in shTNFAIP8 (shT8) knockdown models of human prostate (PC-3, C4-2B), lung (A549, H1299), and pancreatic cancer cell lines (MIAPaC, PANC-1) versus scrambled control cells (scr). Primer sequences used and the anticipated RT-PCR product sizes of various genes were as described earlier [14]. *n* = number of independent qRT-PCR reactions performed
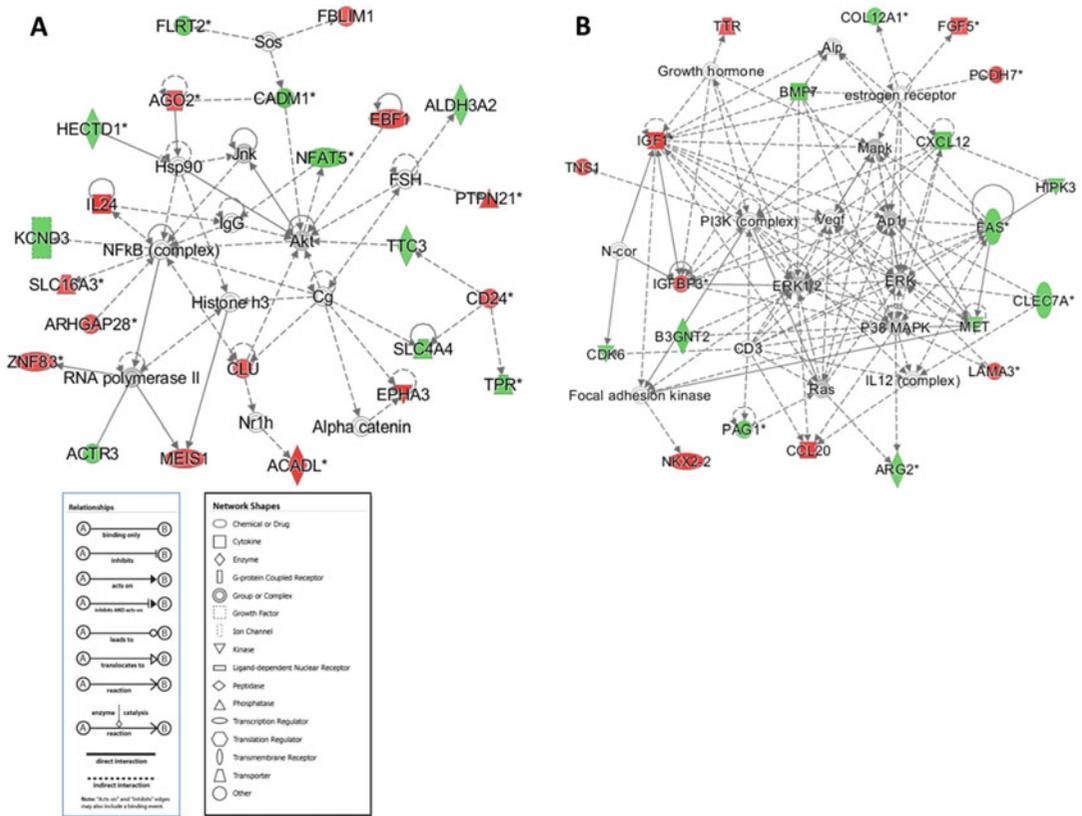
Fig. 6 Transcriptome-based identification of TNFAIP8-centric top gene networks. (**a**) *Top network 1*: Cellular Development, Embryonic Development, Organ Development. (**b**) *Top network 2*: Cancer, Cell Death and Survival, Organismal Injury and Abnormalities. *Green shapes* denote downregulated mRNAs and red shapes indicate upregulated mRNAs in TNFAIP8 knockdown cancer cells. *Blank shapes* (*clear, open*), molecules connected to experimental dataset based on the IPA database showing a strong basis of connection; *Grey shapes*, molecules that but do not meet the designated fold change cutoff (± 1.3); *, duplicates of the molecule seen in the dataset and the one shown in the network was the one with the greatest fold change

Table 4
Top canonical pathways in which TNFAIP8-centric mRNAs and proteins are implicated in cancer cells

Top canonical pathway	Source of data (arrays)	*P*-value
HIF-1α signaling	Antibody	1.58E-10
Glucocorticoid receptor signaling	Antibody	2.09E-10
Leukocyte extravasation signaling	Antibody	6.09E-10
Glioma signaling	Antibody	2.45E-09
Pancreatic adenocarcinoma signaling	Antibody	4.93E-09
Axonal guidance signaling	RNA	2.81E-04
Hepatic fibrosis/hepatic stellate cell activation	RNA	2.07E-03
Clathrin-mediated endocytosis signaling	RNA	2.17E-03
Aryl hydrocarbon receptor signaling	RNA	4.99E-03
Ephrin receptor signaling	RNA	1.06E-02

Table 5
Top diseases and bio functions related to TNFAIP8-centric expression profiles

Diseases and disorders	Arrays (source of data)	*P*-value range
Cancer	Antibody	4.97E-07–2.22E-29
Hematological disease	Antibody	4.97E-07–2.22E-29
Immunological disease	Antibody	4.97E-07–2.22E-29
Organismal injury and abnormalities	Antibody	4.97E-07–2.22E-29
Connective tissue disorders	Antibody	1.14E-07–9.78E-20
Cancer	RNA	4.90E-03–1.03E-08
Gastrointestinal disease	RNA	4.90E-03–1.03E-08
Organismal Injury and abnormalities	RNA	4.90E-03–1.03E-08
Tumor morphology	RNA	4.90E-03–2.18E-06
Renal and urological disease	RNA	4.90E-03–4.48E-06

according to the manufacturer's instructions as we have detailed earlier [13, 23, 24]. Samples were processed in duplicate or triplicate for antibody array analysis. We have observed distinct and overlapping TNFAIP8-centric protein expression profiles in various tumor cell models tested (Table 6). Among the top-ranked genes, TNFAIP8 knockdown correlated with increased expression of the orphan nuclear receptor NR4A1 and alpha 1 adaptin subunit of the adaptor-related protein complex 2 (AP2A1) found in clathrin-coated vesicles in three tumor cell models (Table 6). In addition, TNFAIP8 knockdown correlated with decreased expression of eight proteins (EGFR, PDCL, GTF2F2, IL5, GRAP2, ABL1, AKAP2, GAP43) in at least three tumor cell models (Table 6).

On the PC-3, MDA-MB231, LM2-4175, and MDA-MB435 antibody array datasets, "core" analyses were run with a fold change cutoff of ±1.3. Top canonical pathways, connectivity maps, and top diseases and biofunctions were generated by IPA. Fisher's *P*-values are given as a measure of overlap between observed and predicted gene sets. Based on the integrated proteome analysis, HIF-1α and glucocorticoid signaling were the top two canonical pathways, and cancer was among the top-ranked diseases and biofunctions signifying TNFAIP8-centric changes in cellular proteins (Fig. 7, and Tables 4 and 5).

Table 6
Selected antibody array data showing changes in expression of various proteins in TNFAIP8 knockdown models of prostate cancer (PC-3), breast cancer (MDA-231 and LM2-4175) and melanoma cells (MDA-MB435)

SwissProt	Description (gene symbol)[a]	Fold change in si/shTNFAIP8 vs. Scr (+, high in si/shTNFAIP8; −, low in si/shTNFAIP8)[a]
P17936	Insulin-like growth factor binding protein 3 (IGFBP3)	+1.7 (PC)
Q13596	**Sorting nexin 1 (SNX1)**	**+1.5 (231); +1.5 (LM2)**
P22736	**Nuclear receptor subfamily 4, group A (NR4A1)**	**+1.2 (PC); +1.2 (231); +1.6 (435)**
O95782	**Adaptor-related protein complex 2, alpha 1 (AP2A1)**	**+1.4 (PC); +1.2 (231); +1.4 (435)**
P42336	Phosphoinositide-3-kinase, catalytic (PIK3CA)	−2.6 (LM2)
P41240	c-Src tyrosine kinase (SRC)	−1.7 (LM2)
Q1RMF6	Microtubule-associated protein tau (MAPT)	−1.6 (LM2)
P51955	NIMA (never in mitosis gene a)-related kinase 2 (NEK2)	−1.4 (LM2)
Q9Y3R0	Glutamate receptor interacting protein 1 (GRIP1)	−1.5 (PC)
Q28595	**Microtubule-associated protein tau (MAPT)**	**−1.2 (PC); −1.5 (231)**
P11388	**Topoisomerase (DNA) II alpha 170 kDa (TOP2A)**	**−2.0 (PC); −1.5 (435)**
Q14848	**TNF receptor-associated factor 4 (TRAF4)**	**−2.2 (PC); −1.2 (231)**
P00533	**Epidermal growth factor receptor (EGFR)**	**−1.7(PC); −1.3 (231); −1.3 (435)**
Q13371	**Phosducin-like (PDCL)**	**−1.8 (PC); −1.9 (231); −1.4 (435)**
P13984	**General transcription factor IIF, 30 kDa (GTF2F2)**	**−1.6 (PC); −1.5 (231); −1.2 (435)**
P05113	**Interleukin 5 (colony-stimulating factor) (IL5)**	**−1.5 (PC); −1.2 (231); −1.2 (435)**
O75791	**GRB2-related adaptor protein 2 (GRAP2)**	**−1.7 (PC); −1.2 (231); −1.3 (435)**
Q13691	**v-Abl Abelson murine leukemia viral oncogene (ABL1)**	**−1.5 (PC); −1.4 (231); −3.5 (435)**
Q9Y2D5	**A-kinase (PRKA) anchor protein 2 (AKAP2)**	**−1.5 (PC); −1.4 (231); −1.3 (435)**
P17677	**Growth associated protein 43 (GAP43)**	**−1.4 (PC); −1.3 (231); −1.4 (LM2); −1.5 (435)**

[a]*Bold* indicates that similar changes in protein expression were observed in two or more models of TNFAIP8 knockdown cancer cells

[b]PC-3 (PC), MDA-MB231 (231), and MDA-MB435 (435) cells were treated with siTNFAIP8 or Scr control as described earlier [13]. LM2-4175 cells (LM2) were treated with shTNFAIP8 or Scr control as reported earlier [14]
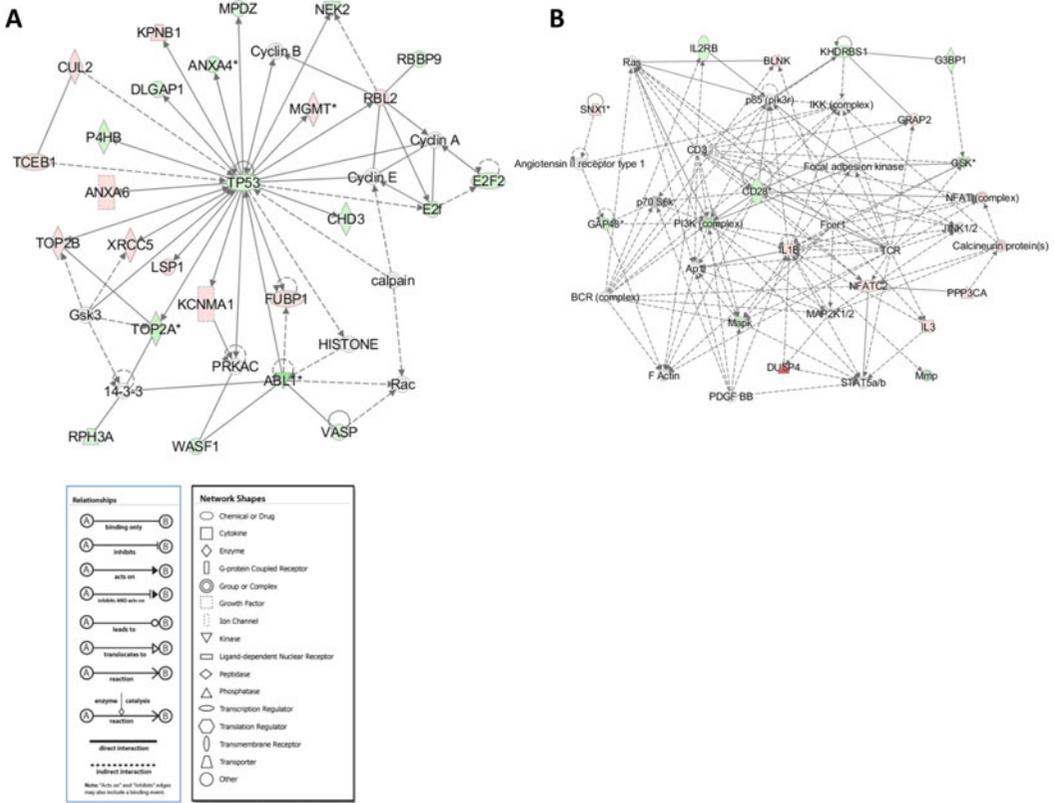
Fig. 7 Proteome-based identification of TNFAIP8-centric top gene networks. (**a**) *Top network 1*: Cancer, Neurological Disease, Organismal Injury and Abnormalities. (**b**) *Top network 2*: Cellular Development, Cellular Growth and Proliferation, Hematological System Development and Function. *Green shapes* denote downregulated proteins and *red shapes* indicate upregulated protein in TNFAIP8 knockdown cancer cells. *Blank shapes* (*clear, open*), molecules connected to experimental dataset based on the IPA database showing a strong basis of connection; *Grey shapes*, molecules that but do not meet the designated fold change cutoff ($\pm$ 1.3); *, duplicates of the molecule seen in the dataset and the one shown in the network was the one with the greatest fold change

4   Conclusions and Future Directions

We have identified a set of at least 13 oncogenic molecules (NFAT5, MALAT1, MET, FOXA1, KRAS, S100P, OSTF1, PIK3CA, SRC, EGFR, IL5, ABL1, and GAP43) and four antiproliferative and apoptotic molecules (IL-24, FAT3, LPHN2 and EPHA3) as TNFAIP8-centric potential regulators of cancer cell survival and progression. Collectively, our data suggest significance of TNFAIP8 expression in the Hypoxia-inducible factor 1α (HIF1-α) signaling pathway and other possible mechanisms of cell survival, proliferation, and invasion. For example, a link between HIF-1α-mediated suppression of fatty acid β-oxidation (FAO) and cancer progression has been reported earlier [25]. HIF-1α seems to inhibit long-chain

acyl-CoA dehydrogenase (LCAD) leading to accumulation of unsaturated fatty acids (UFA) and subsequent inhibition of tumor suppressor PTEN via upregulation of miR-21 [25, 26]. We have observed an increased expression of ACADL in TNFAIP8 knockdown cells. It is conceivable that TNFAIP8 depletion may result in novel activation of FAO and inhibition of tumor growth, in part, by activating the PTEN pathway. Hypoxia-inducible HIF-1α/1β-dependent expression of SLC1A1, gene coding for a glutamate transporter, and enhanced glutamate signaling have been shown to result in activation of SRC kinases and downstream pathways of cell migration and invasion [27]. SLC1A1 expression was found to be decreased in TNFAIP8 knockdown prostate cancer cells. Thus TNFAIP8 may regulate invasiveness in certain cancer cells via SLC1A1 and glutamate receptor-activated SRC kinase pathway.

The orphan nuclear receptor NR4A1 seems to have both pro-apoptotic and growth promoting effects in various cancers [28–32]. More recently, NR4A1-deficient mice which lack nonclassical "patrolling" monocytes (PMo) have been reported to have increased lung metastases [33]. Our data showing increased expression of NR4A1 in TNFAIP8 knockdown tumor cell models may explain earlier observations of decreased pulmonary colonization of melanoma cells in mice systemically treated with TNFAIP8-targeted antisense oligos [12]. Finally, TNFAIP8 knockdown cells showed increased expression of clathrin adaptor-related protein complex 2, alpha 1 subunit (AP2A1). AP2, member of the Assembly Polypeptide family, binds to phosphatidyl inositol (4,5)-bisphosphate (PtdIns4,5P2) in the plasma membrane and is crucial for clathrin-mediated endocytosis (CME) of transmembrane proteins including EGFR [34–38]. Therefore, one speculation is that TNFAIP8 may regulate trafficking of the membrane-bound molecules indirectly by modulating expression levels of the major endocytic clathrin adaptor AP2. Future validation of the key molecules identified in this study is likely to lead to new subset of molecules and functional determinants of cancer cell survival and progression.

Acknowledgements

References

1. Patel S, Wang FH, Whiteside TL, Kasid U (1997) Identification of seven differentially displayed transcripts in human primary and matched metastatic head and neck squamous cell carcinoma cell lines: implications in metastasis and/or radiation response. Oral Oncol 33:197–203

2. Kumar D, Whiteside TL, Kasid U (2000) Identification of a novel tumor necrosis factor-alpha-inducible gene, SCC-S2, containing the consensus sequence of a death effector domain of fas-associated death domain-like interleukin-1beta-converting enzyme-inhibitory protein. J Biol Chem 275:2973–2978

3. Zhang Z, Liang X, Gao L et al (2015) TIPE1 induces apoptosis by negatively regulating Rac1 activation in hepatocellular carcinoma cells. Oncogene 34:2566–2574

4. Sun H, Gong S, Carmody RJ et al (2008) TIPE2, a negative regulator of innate and adaptive immunity that maintains immune homeostasis. Cell 133:415–426

5. Gus-Brautbar Y, Johnson D, Zhang L et al (2012) The anti-inflammatory TIPE2 is an inhibitor of the oncogenic Ras. Mol Cell 45:610–618

6. Fayngerts SA, Wu J, Oxley CL et al (2014) TIPE3 is the transfer protein of lipid second messengers that promote cancer. Cancer Cell 26:465–478

7. Lou Y, Liu S (2011) The TIPE (TNFAIP8) family in inflammation, immunity, and cancer. Mol Immunol 49:4–7

8. Zhang X, Wang J, Fan C et al (2009) Crystal structure of TIPE2 provides insights into immune homeostasis. Nat Struct Mol Biol 16:89–90

9. Horrevoets AJ, Fontijn RD, van Zonneveld AJ et al (1999) Vascular endothelial genes that are responsive to tumor necrosis factor-alpha in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes. Blood 93:3418–3431

10. You Z, Ouyang H, Lopatin D et al (2001) Nuclear factor-kappa B-inducible death effector domain-containing protein suppresses tumor necrosis factor-mediated apoptosis by inhibiting caspase-8 activity. J Biol Chem 276:26398–26404

11. Kumar D, Gokhale P, Broustas C et al (2004) Expression of SCC-S2, an antiapoptotic molecule, correlates with enhanced proliferation and tumorigenicity of MDA-MB 435 cells. Oncogene 23:612–616

12. Zhang C, Chakravarty D, Sakabe I et al (2006) Role of SCC-S2 in experimental metastasis and modulation of VEGFR-2, MMP-1, and MMP-9 expression. Mol Ther 13:947–955

13. Zhang C, Kallakury BV, Ross JS et al (2013) The significance of TNFAIP8 in prostate cancer response to radiation and docetaxel and disease recurrence. Int J Cancer 133:31–42

14. Day TF (2015) Effects of TNFAIP8 knockdown on EGFR and IGF-1R signaling and cytotoxicities of targeted drugs in non-small cell lung cancer cells. Ph.D. Dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University

15. Zhang LJ, Liu X, Gafken PR et al (2009) A chicken ovalbumin upstream promoter transcription factor I (COUP-TFI) complex represses expression of the gene encoding tumor necrosis factor alpha-induced protein 8 (TNFAIP8). J Biol Chem 284:6156–6168

16. Wang Y, Hayakawa J, Long F et al (2005) "Promoter array" studies identify cohorts of genes directly regulated by methylation, copy number change, or transcription factor binding in human cancer cells. Ann N Y Acad Sci 1058:162–185

17. Chakravarty D (2007) Role of SCC-S2 in breast cancer cell proliferation and invasion. Ph.D. Dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of Georgetown University

18. Dong QZ, Zhao Y, Liu Y et al (2010) Overexpression of SCC-S2 correlates with lymph node metastasis and poor prognosis in patients with non-small-cell lung cancer. Cancer Sci 101:1562–1569

19. Liu K, Qin CK, Wang ZY et al (2012) Expression of tumor necrosis factor-alpha-induced protein 8 in pancreas tissues and its correlation with epithelial growth factor receptor levels. Asian Pac J Cancer Prev 13:847–850

20. Duan D, Zhu YQ, Guan LL et al (2014) Upregulation of SCC-S2 in immune cells and tumor tissues of papillary thyroid carcinoma. Tumour Biol 35:4331–4337

21. Eisele L, Klein-Hitpass L, Chatzimanolis N et al (2007) Differential expression of drug-resistance-related genes between sensitive and resistant blasts in acute myeloid leukemia. Acta Haematol 117:8–15

22. Laliberté B, Wilson AM, Nafisi H et al (2010) TNFAIP8: a new effector for Galpha(i) coupling to reduce cell death and induce cell transformation. J Cell Physiol 225:865–874

23. Srivastava M, Eidelman O, Jozwik C et al (2006) Serum proteomic signature for cystic fibrosis using an antibody microarray platform. Mol Genet Metab 87:303–310

24. Jozwik CE, Pollard HB, Srivastava M et al (2012) Antibody microarrays: analysis of cystic fibrosis. Methods Mol Biol 23:179–200

25. Huang D, Li T, Li X et al (2014) HIF-1-mediated suppression of acyl-CoA dehydrogenases and fatty acid oxidation is critical for cancer progression. Cell Rep 8:1930–1942

26. Vinciguerra M, Sgroi A, Veyrat-Durebex C et al (2009) Unsaturated fatty acids inhibit the expression of tumor suppressor phosphatase and tensin homolog (PTEN) via microRNA-21 upregulation in hepatocytes. Hepatology 49:1176–1184

27. Hu H, Takano N, Xiang L et al (2014) Hypoxia-inducible factors enhance glutamate signaling in cancer cells. Oncotarget 5:8853–8868

28. Liu ZG, Smith SW, McLaughlin KA et al (1994) Apoptotic signals delivered through the T-cell receptor of a T-cell hybrid require the immediate-early gene nur77. Nature 367:281–284

29. Lin B, Kolluri SK, Lin F et al (2004) Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor Nur77/TR3. Cell 116:527–540

30. Mullican SE, Zhang S, Konopleva M et al (2007) Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. Nat Med 13:730–735

31. Zhou F, Drabsch Y, Dekker TJA et al (2014) Nuclear receptor NR4A1 promotes breast cancer invasion and metastasis by activating TGF-β signaling. Nat Commun 5:3388. doi:10.1038/ncomms4388

32. Wenzl K, Troppan K, Neumeister P, Deutsch AJ (2015) The nuclear orphan receptor NR4A1 and NR4A3 as tumor suppressors in hematologic neoplasms. Curr Drug Targets 16:38–46

33. Hanna RN, Cekic C, Sag D et al (2015) Patrolling monocytes control tumor metastasis to the lung. Science 350:985–990

34. Höning S, Ricotta D, Krauss M et al (2005) Phosphatidylinositol-(4,5)-bisphosphate regulates sorting signal recognition by the clathrin-associated adaptor complex AP2. Mol Cell 18:519–531

35. Traub LM (2009) Tickets to ride: selecting cargo for clathrin-regulated internalization. Nat Rev Mol Cell Biol 10:583–596

36. McMahon HT, Boucrot E (2011) Molecular mechanism and physiological functions of clathrin-mediated endocytosis. Nat Rev Mol Cell Biol 12:517–533

37. Kelly BT, Graham SC, Liska N et al (2014) AP2 controls clathrin polymerization with a membrane-activated switch. Science 345:459–463

38. Tong J, Taylor P, Moran MF (2014) Proteomic analysis of the epidermal growth factor receptor (EGFR) interactome and post-translational modifications associated with receptor endocytosis in response to EGF and stress. Mol Cell Proteomics 13:1644–1658

39. Shevde LA, Samant RS, Paik JC et al (2006) Osteopontin knockdown suppresses tumorigenicity of human metastatic breast carcinoma, MDA-MB-435. Clin Exp Metastasis 23:123–133

Chapter 8

Network-Oriented Approaches to Anticancer Drug Response

Paola Lecca and Angela Re

Abstract

A long-standing paradigm in drug discovery has been the concept of designing maximally selective drugs to act on individual targets considered to underlie a disease of interest. Nonetheless, although some drugs have proven to be successful, many more potential drugs identified by the "one gene, one drug, one disease" approach have been found to be less effective than expected or to cause notable side effects. Advances in systems biology and high-throughput in-depth genomic profiling technologies along with an analysis of the successful and failed drugs uncovered that the prominent factor to determine drug sensitivity is the intrinsic robustness of the response of biological systems in the face of perturbations. The complexity of the molecular and cellular bases of systems responses to drug interventions has fostered an increased interest in systems-oriented approaches to drug discovery. Consonant with this knowledge of the multifactorial mechanistic basis of drug sensitivity and resistance is the application of network-based approaches for the identification of molecular (multi-)feature signatures associated with desired (multi-)drug phenotypic profiles. This chapter illustrates the principal network analysis and inference techniques which have found application in systems-oriented drug design and considers their benefits and drawbacks in relation to the nature of the data produced by network pharmacology.

Key words Drug–target networks, Drug discovery, Cancer systems biology, Biological network inference, Network vulnerability analysis

1 Introduction

From a systems biology point of view, a network provides an intuitive conceptual representation of the functional and physical relationships among molecular components like genes, proteins, and metabolites that make up life [1, 2]. From a computer science point of view, a network is equated to a data structure whereby data can be organized, visualized, and queried [3]. From a mathematical point of view, a network is formulated through the concepts of graph theory. Network formalism shows a profound diversity in applications scale, nature, and purpose [4, 5]. The approaches developed for network deduction and analysis are highly adaptable to a wide range of

contexts. Conversely, the specification of network node and/or edge attributes, which is at the basis of any network-oriented approach, is highly context-specific.

In this chapter, we illustrate the utility of network-based approaches in anticancer drug discovery surveying strengths and drawbacks of the main applicative cases. The two key challenges facing network applications for drug development are discovering drugs with the expected phenotypic pharmacological profile and identifying the regions of a biological network whose drug-dependent perturbation results in the desired pharmacological outcome. There is widespread evidence that combinations of drugs (*see* **Note 1**) can be more effective than the sum of the effectiveness of the individual drugs [6–8], a result that can be rationalized using principles of modern systems and molecular biology [9–12].

One approach to discover synergistic drug combinations (*see* **Note 2**) is to understand a disease through in silico modeling of the system of interest. Modeling attempts encounter substantial hurdles since they require a precise knowledge of the system being modeled to a level at which it becomes possible to model the actions of perturbations in a predictable way. However, the reliable prediction of complex interactions, such as the synergistic or antagonistic effects that combinations of drugs might exert through cross-pathway wiring, is not yet feasible.

A more pragmatic approach to systematic and synergistic drug discovery combines high-throughput screening of drugs with genotypic profiles and seeks to detect combinations of genotypic features which are associated with drug profiles in a disease model of interest. In this direction, inference of drug–target interaction networks has proven a valuable technique to link the drug activity to the disease-relevant molecular context [13–15].

Equally important, network analysis at the level of individual nodes offers a number of benefits. Network node centrality metrics can be used for characterizing candidate drug targets, for instance, through the prioritization of predictive molecular features based on the specificity and sensitivity of their response to multiple drugs, in cancer-specific or cross-cancer settings. Furthermore, several compelling theoretical and experimental studies support the notion that the topological properties of the network nodes help to enlighten the system function of the molecular components they represent [16–18]. For instance, a relationship has been proposed between the essentiality of individual components in a biological system and several indices of connectivity centrality of the corresponding nodes in the network which abstracts the biological system [17, 19, 20]. Target essentiality has been shown to be related to drug side effects and, therefore, should be taken into consideration in rational drug design [21, 22]. Of special interest are the attempts to characterize network nodes by the size of the effects caused by individual nodes on network stability in response to a

perturbation [23, 24]. A recent characterization of essential genes by topological properties in a perturbation sensitivity network has proven the capability of several network topological node properties to discriminate between essential and nonessential genes [20]. Furthermore, novel centrality metrics have been specifically introduced for the purpose of gauging network dynamical properties at the node level. Therefore, network analysis certainly is a valuable tool for target identification in drug development.

In this chapter, we show that network-based approaches offer a versatile set of tools to computationally support drug discovery programs. Network analysis supports the characterization of candidate targets by virtue of their position in the disease network as well as of the feasibility of their modulation to achieve a beneficial pharmacological outcome. Integration of network inference approaches with combinatorial drug pharmacology holds the promise to meet the pressing clinical need for robust molecular correlates of anticancer drug response.

2   Network-Based Approaches in Drug Targets Identification

Drug discovery seeks to identify small molecules that potently and selectively modulate disease-relevant functions of target proteins. There are two fundamental approaches to understanding the action of small molecules on biological systems. Target-based approaches (reverse chemical genetics) begin with target validation, which is a time-consuming process that involves demonstrating the relevance of the protein for the diseases of interest, and proceed with a biochemical assay to find out candidate drugs affecting the target. It is presumed that drugs affecting the target would affect the desired phenotype; however, this impact needs to be characterized by mechanism-of-action studies. Over the past decade, advances in assay technology have increasingly facilitated the rapid expansion of phenotype-based approaches. Phenotype-based approaches (forward chemical genetics) begin with the identification of a disease-relevant phenotype in a model system, which is accompanied by a screen where candidate drugs are tested for their impact on the phenotype. Candidates must then undergo target identification, followed by mechanism-of-action studies, to determine the protein or proteins responsible for the observed phenotypic change, which can be a complex endeavor. In summary, both target- and phenotype-based approaches benefit from pathway or network information to gain insight into the drug-dependent mechanisms eliciting the desired biological response.

Historically, the "one gene, one drug, one disease" paradigm assumed that a single direct interaction between a drug and a target was responsible for phenotypic observations. However, the limitations of such an approach were clearly shown by the rates of

late-stage drug failure in clinical development which are principally due to lower than desired efficacy and/or to clinical safety or toxicology [25]. A fundamental factor of the success of a drug is how the robustness and fragility, which are intrinsic to biological systems, are exploited in terms of disease onset and progression. Robustness permits biological systems to maintain their characteristic behavior in response to internal or external perturbations [9, 10, 26] and depends on: (1) system control which introduces regulatory loops to ensure system homeostasis and stability; (2) failsafe mechanisms which, by means of compensatory pathways, enable a system to keep functioning when a component is disabled; and (3) modularity which prevents perturbations from spreading through the whole system. Biological robustness is valuable to interpret drug efficacy and side effects. A drug is effective when it hits the point of fragility but can be ineffective when system robustness compensates for any change caused by the drug, through genetic diversity [27–29] or feedback loops [30]. Drug side effects can result from the interference with an unexpected point of fragility of the system [31, 32]. The awareness of biological system robustness prompted the development of combinatorial drug therapy which leverages the effects of multiple drugs to get maximal overall effect [6, 33]. There are several reasons underlying the success of molecularly targeted cancer therapeutics based on drug combinations. For instance, the administration of multiple drugs could disrupt compensatory mechanisms or the combination of multiple drugs could be designed in such a way that a drug targets the point of fragility induced by the other drugs in the combination [34–36]. Given these considerations, drug design clearly requires network-oriented approaches to improve drug efficacy and selectivity by identifying drugs affecting multiple oncogenic signal transduction pathways and drug combinations [37, 38]. In addition to the above theoretical reasons, network-based approaches are advantageous for evaluating and prioritizing the most promising potential drug combinations among the sheer number of mathematically possible ones.

Network topology analysis has offered important clues on drug target hypotheses as well as on the characterization of network topological properties of confirmed targets. Network inference approaches have permitted to formulate mechanistic hypotheses, rather than targets per se. Furthermore, it is worth to note that identifying drug targets and understanding their mechanisms of action actually proceed through a combination of network approaches, along with cell- or organism-based phenotypic assays, to test increasingly specific target hypotheses.

The assumption that network topology could provide valuable information to identify candidate drug targets dates back to 15 years ago when early network analysis indicated the possibility of a direct correlation between the essentiality and the degree of

connectivity of nodes [17]. Subsequent reanalysis of the data challenged this view, according to which more highly connected hubs in a network are more likely to be essential [39]. The relationship between network topology and the system function of a protein when it is perturbed was refined by focusing on metrics such as the betweenness centrality [40] and the bridging centrality [41] in addition to the degree one. Interestingly, an analysis of centrality was conducted also on a human network of the interactions between all US-approved drugs and therapies and showed prominent role for drugs with high betweenness centrality values [42].

If we are to exploit network analysis to identify targets and to predict likely outcomes of their perturbations, then a closer model of drug action on a target has to consider the drug-dependent network dynamics, by capturing the target response to changes in its immediate neighborhood as well as the propagation to distant nodes [43, 44]. To understand the system function response to a perturbation, responsive mathematical models have to account for the interplay between network topology and network dynamics [23, 45]. In this respect, lately a new centrality metrics, named vibrational centrality [46], was borrowed from physics and reformulated in the complex network context permitting to shift the attention from static topological properties to dynamical properties of a perturbed network at the node level. At higher level of mechanistic detail, dynamical models based on the law of mass action have been successfully applied to quantitatively describe the effects of chemical perturbations, used either in isolation or in combination, in several signaling pathways [47, 48]. The main issues are related to the identification of the symbolic model, usually in the form of ordinary reaction rate equations, and to the estimation of the parameters [49, 50]. In the context of phenotypic drug screens, these equations are difficult to set up and parameterize since the analytical form of the dynamics is largely unknown and the estimation of parameters is not approachable. Consequently, differential equation-based models are more useful in the late-stage refinements of the mechanism-of-action of candidate drugs than in the profiling of drug action [51].

Logic-based models permit to bypass some hurdles of mass action-based models like the requirement for parameter specification [52]. Here, interactions are not interpreted as direct physical interactions but as measures of influence between network nodes. Functions of discreet logic are used to connect nodes through gates and to infer functions best capturing the observed dynamics of the data. These approaches have been applied to interpret drug response in triple-negative breast cancer and to suggest therapeutics [53–55].

Phenotype-based drug profiling does not usually grasp the mechanistic aspects of drug action and therefore are not easily tractable by previous approaches. More often, the readouts of individual assays or, more recently, of multiple complementary assays could be

powerfully analyzed by statistical association-based approaches. Bayesian network inference is a promising tool since it can illustrate the influences of nodes on each other by probabilistic dependence relationships [56]. Bayesian models can be deduced from experimental data provided that the a priori probability distributions on the data are supplied. Advantages of Bayesian models include the possibility to accommodate noise inherent in the data in a probabilistic framework which permits to estimate credibility intervals of the parameters associated with the inferred relationships. However, owing to the probabilistic nature of the Bayesian modeling approach, the inference of statistically significant dependencies requires many simultaneous observations of tested molecules. Another caveat, which is of practical importance in drug discovery, is the restriction of inferred networks to be acyclic whereas system responses to perturbations are known to contain feedback loops. Finally, it is worth to note that the probabilistic framework, which could be powerful as we acknowledged, could turn in a severe drawback since in drug screening the a priori knowledge on the data is incomplete and could lead to inappropriate selection of the a priori probability distributions on the data/parameters.

The limited knowledge on a priori data distributions characterizes many drug discovery programs which usually rely on systematic high-throughput and unbiased data. First, a drug screen is conducted to identify effective drug combinations; second, a high-throughput profiling of one or more types of genomic features is conducted to translate genotype features into markers of drug sensitivity or resistance [14, 57, 58]. Therefore, the raw data delivered by those approaches consist of data on drug activity and of data from genotypic profiling. In sophisticated experiments, drug activities could be described by multiple parameters and genotype data could consist of multidimensional (DNA copy number, mRNA abundance, protein abundance) genomic profiles. The next step is to associate the drug activity profile with genotypic changes to obtain association profiles between drugs and genotypic features. The ultimate goal is to select effective drugs or drug combinations and to obtain a multi-feature genomic signature of drug(s) response (Fig. 1). Non-model based approaches for network inference are therefore particularly demanded.

Fig. 1 (continued) (**b**) Various computational techniques provide a matrix of the association between each drug activity and each genotypic feature. (**c**) A combination of multiple genotypic features associated with the response of a drug across the samples is finally deduced. For simplicity, the plot illustrates the multi-feature signature for a single drug response. The heat maps display the values measured for the genotypic features which are included in the signature across the samples (*blue* corresponds to lower values, *red* to higher ones). Features are separated depending on whether they are associated with drug sensitivity or resistance. To the *left* of each feature is a *bar* indicating the effect size. Bars in *yellow* are negative effects, indicating features associated with drug sensitivity, and bars in *purple* are positive effects, indicating features associated with drug resistance. Drug response estimates across samples are shown at the bottom. For clarity, only the top features associated with sensitivity and resistance are shown
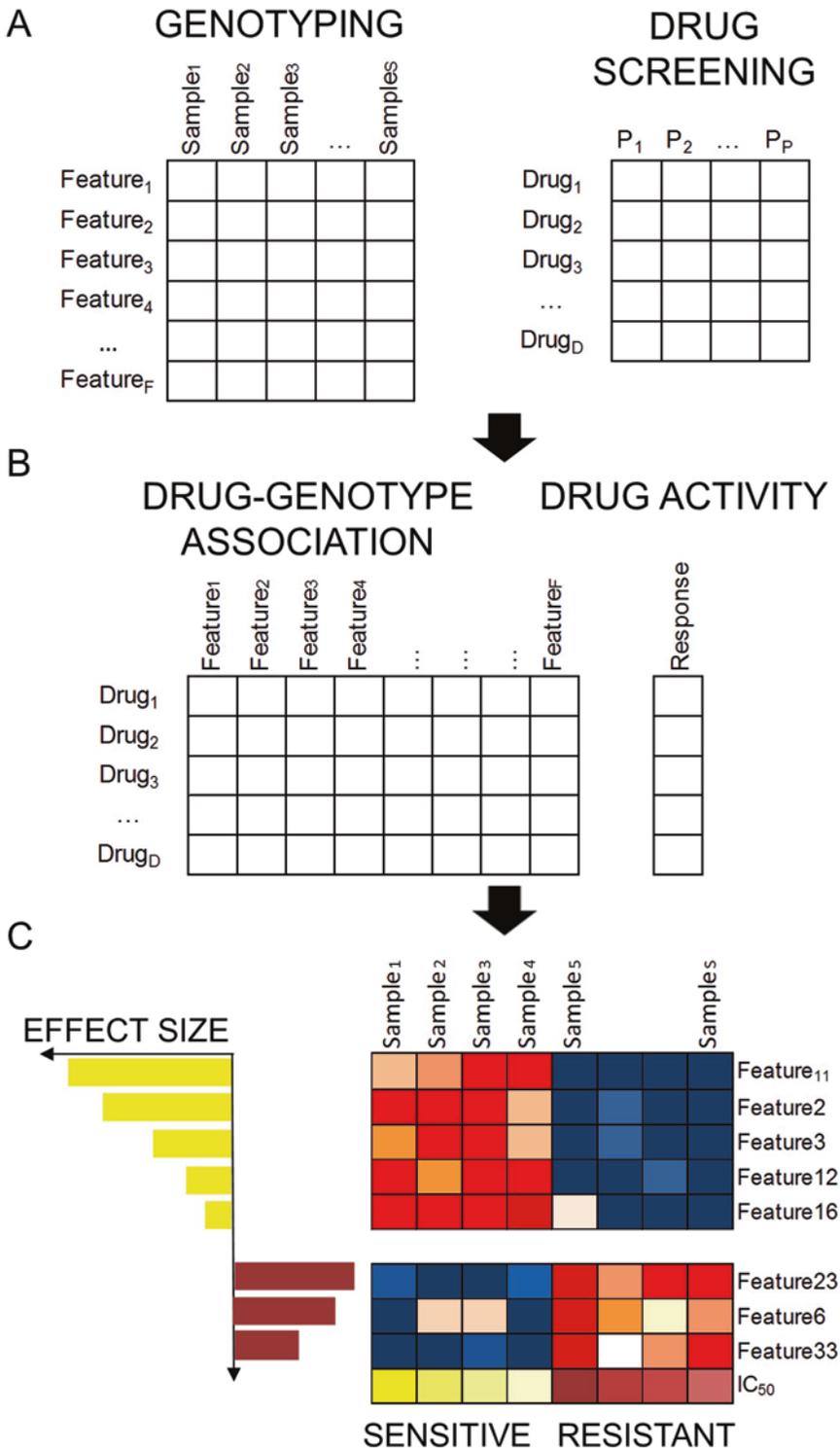
Fig. 1 Multi-feature signature of drug response. (**a**) A panel of tumor samples $\{S_1,\cdots,S_S\}$ is characterized by multiple genomic features and a drug screen quantifies the effect of drug treatment on the phenotype of interest across the same samples. Drug activity can be described by a single parameter or by multiple parameters $\{P_1\cdots P_P\}$ set (e.g., half-maximal inhibitory concentration (IC_{50}) or slope of the dose–response curve).

3   Methods

3.1   Non-Model Based Inference of Drug–Target Networks

Computational techniques can rely on the combination between phenotype-based drug profiling and genotype-based profiling. Drug treatment efficacy is usually scored by response parameters such as half-maximal inhibitory concentration (IC_{50}), i.e., the drug concentration at which the cancer cell growth is inhibited by 50 %, or the area under the dose–response curve, or by the drug sensitivity score given by the closed-form integration of the area under the estimated dose–response curve.

Technical advances in genomic sequencing and genotyping have enabled deeper characterization of drug-profiled samples by high-throughput measurements across multiple types of genetic features. Association profiles are then quantified between the phenotypic profiles of N drugs $(D_1, D_2, \ldots, D_N)$ and the genotypic profiles of P potential targets $\boldsymbol{T} = (T_1,,,T_2,,,\ldots,,,T_P)$.

The aim is to infer efficient individual drugs or combination drugs and the sets of genotypic features that best report on drugs action. The approaches developed by Pal et al. [59] and Tang et al. [60] are suitable to represent the data-driven approaches which are currently in use to identify drug targets from the genomic characterization of phenotypic drug profiling. Here, we do not illustrate these approaches in details, but we review their general workflow explaining the general mathematical formalization of the inference and its solution. In general terms, the inferential procedure first selects efficient drugs or combination drugs and then identifies the signature of targets best predicting the phenotype measured in response to the selected drugs. In the first step, drugs are selected through effect-based approaches or dose-effect-based approaches [8]. The second step implements a minimization problem, where the objective function is the prediction error of the efficacy of the subset of most efficient drugs $D^* \subseteq D$. In Tang et al. [60], it is defined as follows:

$$Err(E^{D_{new}} \left| D_{new}(T^*) \equiv \left| E(D_{new} | T^*) - \hat{E}(D_{new}) \right|$$

where D_{new} is the new drug combination in the testing set, T^* is the set of potential targets of D_{new}, $E(D_{new} | T^*)$ is the predicted efficacy of D_{new} on T^*, and $E(D_{new})$ is the observed efficacy of D_{new}. $E(D_{new} | T^*)$ is directly estimated from the input data as follows:

$$E(D_{new} | T^*) \equiv \frac{\sum_{i \in N} I(D_i^* \subset D_{new}) \hat{E}_i}{\sum_{i,j \in N} I(D_i^* \subset D_{new}) \hat{E}_i + I(D_{new} \subset D_j^*)(1 - \hat{E}_j)}$$

where $I(\cdot)$ is an indicator function, equal to one when the argument is true, and zero otherwise. In particular, $E\left(D_{new} \mid T^{*}\right)=1$ if D_{new} is a superset of $D^{\star}$ and $\hat{E}_{i}=1$, whereas $E\left(D_{new} \mid T^{*}\right)=0$ if D_{new} is a subset of $D^{\star}$ and $E_{i}=0$.

A simple data-driven procedure to infer T^{*} from the initial set of potential targets T is a procedure that constructs subsets of T and evaluates the error in the prediction of drug efficacy for each of these subsets. In Tang et al. [60]) T^{*} is selected by minimizing an average leave-one-out error (LOO):

$$\frac{1}{N}\sum_{i \in N}\mathrm{Err}\left(T_{i} \mid D_{i}^{(T_{-i})}\right).$$

The overall picture of this inferential scheme is shown in Fig. 2.

3.2   Network Vulnerability Analysis: Vibrational Centrality

3.2.1   Vibrational Centrality

Networked system response to perturbations can be independent of the organizational architecture of the network. To measure the susceptibility of network nodes to external changes (i.e., drugs), a novel centrality measure, termed *vibrational centrality* proposed by E. Estrada [46], is appropriate. The definition of this index is based on the analogy of a network with a physical system in which the nodes are rigid spheres and the edges are springs. The effects of stresses on such a system are modeled as nodes displacements, i.e., nodes reactions to changed conditions could be represented in terms of nodes vibrations, that are deviations of nodes positions from their equilibrium ones.

In order to define the position of nodes, the network is embedded into an n-dimensional Euclidean space (n being the number of nodes in the network) represented by the Moore–Penrose pseudo-inverse of graph Laplacian $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where D is the diagonal matrix of degrees and A is the graph adjacency matrix. Henceforth we denote by $\boldsymbol{L}^{+}$ the pseudo-inverse of L. Each diagonal entry of $\boldsymbol{L}^{+}$, denoted as l_{ii}^{+} for the i-th node, represents the squared distance of node i to the origin in this n-dimensional space and provides a measure of the node's topological centrality, given as

$$C(i) = \frac{1}{l_{ii}^{+}}. \tag{1}$$

Closer the node i is to the origin in this space, or equivalently lower the l_{ii}^{+}, more topologically central node i is [61].

A network subjected to some sort of external stresses can be modeled by a network submerged in a thermal bath at temperature T. This abstraction allows to define the vibrational potential energy of the network as it is usually defined for a system of microscopic particles, i.e.,

Input data

Drug set: $D = (D_1, D_2, ..., D_N)$

Potential target set: $T = (T_1, T_2, ..., T_P)$

D	T_1	T_2	T_3	...	...	...	T_P	Efficacy
D_1	1	0	1	...	...	...	...	E_1
D_2	0	0	1	...	...	...	...	E_2
D_3	0	1	1	...	...	...	...	E_3
...	0	1	0	...	...	...	...	...
...	1	1	0	...	...	...	...	...
...	1	1	0	...	...	...	...	...
D_N	0	0	1	...	...	...	...	E_N

Step 1

Selection of the most efficient drugs subset

$$D^* \subseteq D$$

D	T_1	T_2	T_3	...	...	...	...	T_P
D_1	1	0	1	...	...	...	...	...
D_3	0	1	1	...	...	...	...	...
D_N	0	0	1	...	...	...	...	...

Step 2

Construction of the combinatorial space of potential targets (e.g. through leave-one-out method)

$$(D^*, T_{-1})$$
$$(D^*, T_{-2})$$
$$(D^*, T_{-3})$$
$$\cdots$$
$$(D^*, T_{-P})$$
$$T_{-i} \equiv \{T_k\}_{k \neq i}, \quad k = 1, 2, \ldots, P$$

Step 3

Selection of T_{-i} minimizing the efficacy predition error for D^*:

$$\frac{1}{N} \sum_{i \in N} \mathrm{Err}(T_i | D_i^{(T_{-i})})$$

Output

D_1 —⊣ T_3

D_1 ⊣ T_1

T_3 ⟵ D_3

D_3 ⟶ T_4

Example: $D^* = \{D_1, D_3\}$

$P = 4$ $\qquad T_{-2} = \{T_1, T_3, T_4\}$

Fig. 2 Scheme of non-model based inference of drug–target network. The inference procedure reported earlier [60] consists of three steps. *Step 1* selects the most efficient drug, and *steps 2* and *3* deal with drug–target network model construction

$$V(x) = \frac{k}{2} x^\top L x \tag{2}$$

where k is the spring constant, and x is the vector whose i-th entry is the displacement x_i of the i-th node. The probability distribution of the displacemen of the nodes is given by the Boltzman distribution

$$P(x) = \frac{e^{-\frac{1}{T}V(x)}}{Z} = \frac{1}{Z} \exp\left(-\frac{k}{2T} x^T L x\right) \tag{3}$$

where the partition function Z of the network is defined as follows

$$Z \equiv \int dx \exp\left(-\frac{k}{2T} x^T L x\right).$$

Given $P(\mathrm{x})$, the mean displacement of a node i, is by definition

$$\langle \Delta x_i \rangle \equiv \sqrt{\int x_i^2 P(x) dx} \tag{4}$$

It can be shown [62] that

$$\langle \Delta x_i \rangle = \sqrt{\frac{T}{k}(L)_{ii}} \tag{5}$$

Equations (2) and (3) describe the potential vibrational energy of a network submerged into a thermal bath at the temperature T.

Vibrational centrality was found to achieve high resolution in the identification of the most vulnerable nodes in a network, compared to standard topological measures such as node degree, which can account for only the nearest-neighbors of a node to quantify the relative contribution of the node to network integrity [46].

4   Example of Network Centrality Characterization of Cancer Genes for Drug Discovery

We set out to explore the relationship between the effect of cancer-related gene changes on drug response and the centrality of the protein products in a network of protein interconnections. The Genomics of Drug Sensitivity in Cancer (GDSC) project is an academic research program to identify molecular features of cancers that predict response to anticancer drugs [63]. Data about gene effects on the sensitivity of a rich panel of cancer cell lines to a wide selection of compounds representing approved drugs or drugs under development were downloaded from the GDSC database (www.cancerRxgene.org). In fact, this resource

provides the results of an analysis of the association of gene mutational state (gene somatic mutation, amplification, or deletion), as reported in the Catalogue of Somatic Mutations in Cancer database, with cell line drug sensitivity data, as quantified by the IC_{50} value and the slope of the dose–response curve. The collected information on drug–gene association was then framed in a bipartite weighted network where the weight of a drug–gene edge represented the inverse of the statistical significance of the multivariate analysis of variance (MANOVA) test. We then mapped the proteins encoded by the genes in the drug–gene association network on the IntAct protein interaction network through the genes2networks tool (http://amp.pharm.mssm.edu/genes2networks/). An analysis of node centrality was conducted both in the drug–gene association network and in the protein interaction network. Centrality of gene nodes in the drug–gene association network was quantified by the sum of their edges weights (node strength). In this way, according to the definition of drug–gene edge weights, the nodes characterized by the highest strength values in the drug–gene association network correspond to the strongest modifiers of drug response. The centrality analysis of the proteins encoded by the screened genes in the protein interaction network was based on the degree, betweenness and vibrational centrality metrics. We then conducted an analysis of the relationship between node centrality in the protein interaction network, as assessed by the mentioned centrality metrics, and the node strength in the drug–gene association network (Fig. **3**). This analysis suggested that node degree and betweenness in the protein interaction network were not informative on node centrality in the drug–gene association network since degree and betweenness values in the protein interaction network were found to be low, irrespectively of the effect of the corresponding genes on drug sensitivity. Instead, the assessment of node centrality by the vibrational centrality metric identified an interesting pattern since differences in node vibrational centrality in the protein interaction network seemed to correspond to differences in node strength in the drug–gene association network. Interestingly, the strongest nodes in the drug–gene association network were found to show the lowest vibrational centrality in the protein interaction network, i.e., to be particularly stable nodes in the network. Considering also the previous observation on low betweenness values, this analysis suggests that the strongest modifiers of drug response tend to be connected by particularly stable edges and, nonetheless, to only moderately contribute to connect nodes to each other in the protein interaction network. We present this example study to introduce a way of generating hypotheses on the attractiveness of proteins with regard to the discovery of drug modulators through network centrality analysis.
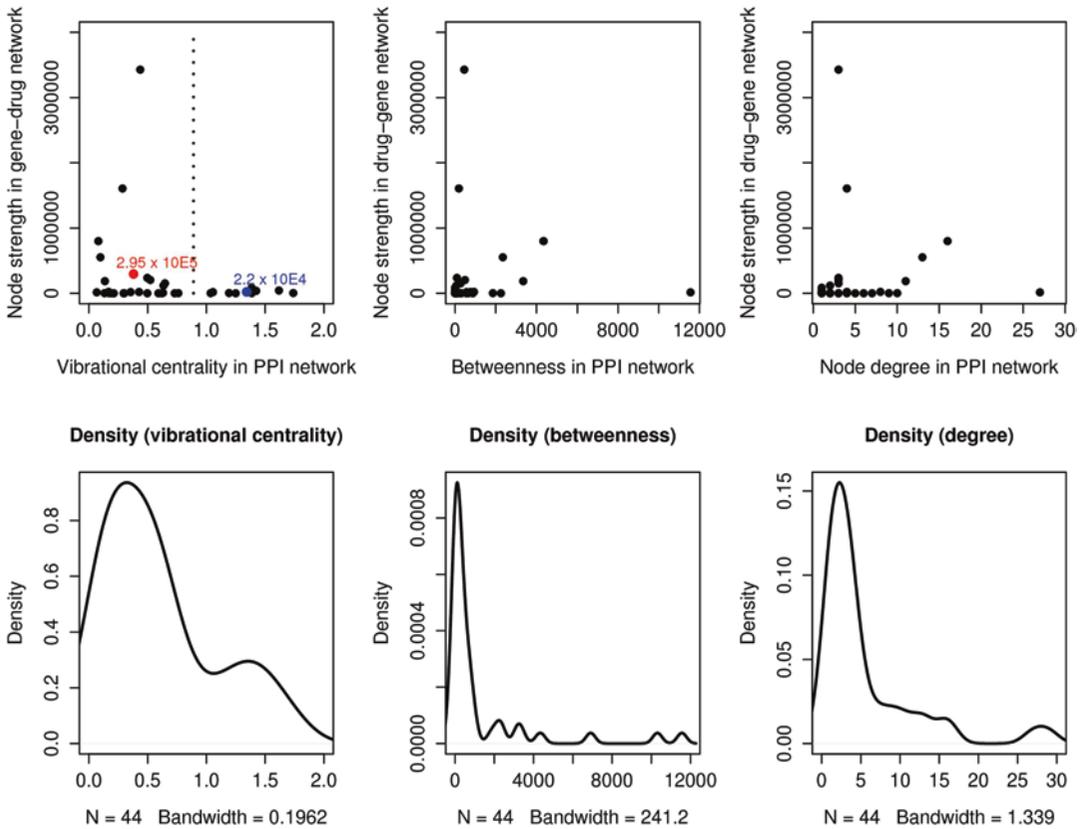
Fig. 3 Node strength in drug–gene network versus node centrality in protein interaction (PPI) network. The strength of a node is the sum of the weights of its incident edges. In the drug–gene network considered in this study the edge weight is the inverse of the *p*-value scoring the association of the mutational state of a gene with the drug response. The density plots of the centrality measures of the PPI network show that only the vibrational centrality values are partitioned into two intervals identified by the width of the corresponding peaks. The higher peak covers values of vibrational centrality smaller than those underlying the second peak. The partition of vibrational centrality values in the protein interaction network (indicated by the *dotted line* in the first plot) corresponds to a partition of node strength values in the drug–gene association network: the *red* and the *green* points in the first plot identify the average value of node strength corresponding to the two groups partitioning the values of vibrational centrality. We see that the average node strength of the nodes having vibrational centrality smaller that 1 is one order of magnitude greater than then average node strength of the nodes with vibrational centrality greater than 1 (2.95×10^5, and 2.20×10^4, respectively). In the most common behavior, the strongest nodes in the drug–gene association network have lowest vibrational centrality in the PPI network. Additionally, the betweenness (an degree) in PPI networks is low for the great majority of the nodes

5   Conclusions

Network-oriented approaches are key components in rational drug design for several theoretical and pragmatic reasons. Theoretically, the major endeavor is to account for the complexity of the multifactorial basis of the response of biological systems to drug interventions. Since biological systems have evolved to be

robust in the face of internal and external perturbations, they can also show robust against drug intervention. Network-based approaches provide a way of thinking about the systemic aspects of drug sensitivity and resistance and therefore are expected to valuably complement screening-based approaches. Indeed, a systematic and efficient network/based formalization of data gathered by combining phenotype/based drug screens with genome/wide genotypic profiles has proven to be useful for inferring the best combination of molecular cognates of the desired pharmacological output of single drugs or of multicomponent drugs.

From a practical point of view, the notion of biological systems as networked systems has led to the consideration that the realization of the full potential of molecularly targeted cancer therapeutics depends on identifying and optimizing the action of combinations of drugs rather than of individual drugs. This adds a considerable level of combinatorial complexity in the drug discovery and therapy design process. It is practically prohibitive to evaluate the total number of possible combinations of available drugs and the corresponding targets. Network representation of experimental information allows an in silico exploration of the combinatorial space.

Network analysis can be used not only to optimize drug efficacy by the inference of the optimal drug target combinatorics but also to assess drug specificity; for instance, mapping the targets found to be sensitive to a drug or drug combination onto a protein interaction network can increment our understanding of drug side effects thus aiding informed decision-making. Furthermore, analyzing the protein interaction network context of chemically tractable targets can help pinpoint potential alternative but previously neglected subnetworks in future drug discovery initiatives. In conclusion, the achievement of positive outcomes by modern systems-oriented drug design is crucially dependent on the combination of systematic drug screening, genotyping profiling with network-based approaches.

6 Notes

1. *Drug combination.* An optimized combination of multiple drugs including syncretic or congruous drugs. A combination of drugs is syncretic if it is composed of two or more drugs, at least one of which is not used individually to treat the target disease. A combination of drugs is congruous if each drug has been individually used to treat the target disease.

2. *Synergy and antagonism.* They represent respectively greater or lesser effects for drugs in combination than the simple additive effect expected from the knowledge of the effects of

each drug individually. Useful analyses of mathematical and pharmacological concepts underlying the assessment of the effects of combinatorial signals (e.g., drugs) are provided in refs. [8] and [33, 64].

References

1. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664

2. Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113

3. Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms, 2nd edn. MIT Press and McGraw-Hill Book Company, Cambridge, MA

4. Kitano H (2001) Exploring complex networks. Nature 410:268–276

5. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435:814–818

6. Al-Lazikani B, Banerji U, Workman P (2012) Combinatorial drug therapy for cancer in the post-genomic era. Nat Biotechnol 30:679–692

7. Keith CT, Borisy AA, Stockwell BR (2005) Multicomponent therapeutics for networked systems. Nat Rev Drug Discov 4:71–78

8. Foucquier J, Guedj M (2015) Analysis of drug combinations: current methodological landscape. Pharmacol Res Perspect 3:e00149

9. Kitano H (2004) Biological robustness. Nat Rev Genet 5:826–837

10. Stelling J, Sauer U, Szallasi Z et al (2004) Robustness of cellular functions. Cell 118: 675–685

11. Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461

12. Vidal M, Cusick ME, Barabási AL (2011) Interactome networks and human disease. Cell 144:986–998

13. Covell DG (2015) Data mining approaches for genomic biomarker development: applications using drug screening data from the cancer genome project and the cancer cell line encyclopedia. PLoS One 10, e0127433

14. Lamb J, Crawford ED, Peck D et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313:1929–1935

15. Woo JH, Shimoni Y, Yang WS et al (2015) Elucidating compound mechanism of action by network perturbation analysis. Cell 162:441–451

16. Zhu M, Gao L, Li X et al (2009) The analysis of the drug-targets based on the topological properties in the human protein-protein interaction network. J Drug Target 17:524–532

17. Jeong H, Mason SP, Barabśi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411:41–42

18. Khuri S, Wuchty S (2012) Essentiality and centrality in protein interaction networks revisited. BMC Bioinformatics 16:109

19. Brush ER, Krakauer DC, Flack JC (2013) A family of algorithms for computing consensus about node state from network data. PLoS Comput Biol 9:e1003109

20. Yang L, Wang J, Wang H et al (2014) Characterization of essential genes by topological properties in the perturbation sensitivity network. Biochem Biophys Res Commun 448:473–479

21. Wang X, Thijssen B, Yu H (2013) Target essentiality and centrality characterize drug side effects. PLoS Comput Biol 9:e1003119

22. Peng Q, Schork NJ (2014) Utility of network integrity methods in therapeutic target identification. Front Genet 5:12

23. Barzel B, Barabási AL (2013) Universality in network dynamics. Nat Phys 9

24. Estrada E (2012) The structure of complex networks. Theory and applications, 1st edn. Oxford University Press, Oxford, UK

25. Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Discov Today 10:139–147

26. Kitano H (2010) Violations of robustness trade-offs. Mol Syst Biol 6

27. Kassouf W, Dinney CP, Brown G et al (2005) Uncoupling between epidermal growth factor receptor and downstream signals defines resistance to the antiproliferative effect of gefitinib in bladder cancer cells. Cancer Res 65(10):524–535

28. Sergina NV, Rausch M, Wang D et al (2007) Escape from her-family tyrosine kinase inhibitor therapy by the kinase-inactive her3. Nature 445:437–441

29. Meng J, Peng H, Dai B et al (2009) High level of akt activity is associated with resistance to mek inhibitor azd6244 (arry-142886). Cancer Biol Ther 8:2073–2080

30. Rodrik-Outmezguine VS, Chandarlapaty S, Pagano NC et al (2011) mTOR kinase inhibition

causes feedback-dependent biphasic regulation of akt signaling. Cancer Discov 1:248–259

31. FitzGerald GA, Patrono C (2001) The coxibs, selective inhibitors of cyclooxygenase-2. N Engl J Med 345:433–442

32. Ito T, Ando H, Suzuki T et al (2010) Identification of a primary target of thalidomide teratogenicity. Science 327:1345–1350

33. Jia J, Zhu F, Ma X et al (2009) Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov 8:111–128

34. Modi S, Stopeck A, Linden H et al (2011) Hsp90 inhibition is effective in breast cancer: a phase ii trial of tanespimycin (17-aag) plus trastuzumab in patients with her2-positive metastatic breast cancer progressing on trastuzumab. Clin Cancer Res 17:5132–5139

35. Nahta R, Hung MC, Esteva FJ (2004) The her-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells. Cancer Res 64:2343–2346

36. Meng J, Dai B, Fang B et al (2010) Combination treatment with mek and akt inhibitors is more effective than each drug alone in human non-small cell lung cancer in vitro and in vivo. Cell 5:e14124

37. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4:682–690

38. Pe'er D, Hacohen N (2011) Principles and strategies for developing network models in cancer. Cell 144:864–873

39. Zotenko E, Mestre J, O'Leary DP et al (2008) Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Comput Biol 4:e1000140

40. Yu H, Kim PM, Sprecher E et al (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3:e59

41. Hwang WC, Zhang A, Ramanathan M (2008) Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. Clin Pharmacol Ther 84:563–572

42. Nacher JC, Schwartz JM (2008) A global view of drug-therapy interactions. BMC Pharmacol 8

43. Almaas E, Kovács B, Vicsek T et al (2004) Global organization of metabolic fluxes in the bacterium Escherichia coli. Nature 427:839–843

44. Csermely P (2004) Strong links are important, but weak links stabilize them. Trends Biochem Sci 29:331–334

45. Zhang X, Zhang Z, Zhao H et al (2014) Extracting the globally and locally adaptive backbone of complex networks. PLoS One 9:e100428

46. Estrada E, Hatano N (2010) A vibrational approach to node centrality and vulnerability in complex networks. Physica A 389:3648–3660

47. Birtwistle MR, Hatakeyama M, Yumoto N et al (2007) Ligand-dependent responses of the erbb signaling network: experimental and modeling analyses. Mol Syst Biol 3:144

48. Iadevaia S, Lu Y, Morales FC et al (2010) Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. Cancer Res 70:6704–6714

49. Chou IC, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219:57–83

50. Liu Y, Gunawan R (2014) Parameter estimation of dynamic biological network models using integrated fluxes. BMC Syst Biol 8:127

51. Faratian D, Goltsov A, Lebedeva G et al (2009) Systems biology reveals new strategies for personalizing cancer medicine and confirms the role of pten in resistance to trastuzumab. Cancer Res 69:6713–6720

52. Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. Biochemistry 49:3216–3224

53. Lee MJ, Ye AS, Gardino AK et al (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell 149:780–794

54. Sahin O, Fröhlich H, Löbke C et al (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. BMC Syst Biol 3:1

55. Aldridge BB, Saez-Rodriguez J, Muhlich JL et al (2009) Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. PLoS Comput Biol 5:e1000340

56. Sachs K, Perez O, Pe'er D et al (2005) Causal protein-signaling networks derived from multiparameter single-cell data. Science 308:523–529

57. Barretina J, Caponigro G, Stransky N et al (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483:603–607

58. Shoemaker RH (2006) The nci60 human tumour cell line anticancer drug screen. Nat Rev Cancer 6:813–823

59. Pal R, Berlow N (2012) A kinase inhibition map approach for tumor sensitivity prediction and combination therapy design for targeted drugs. Pac Symp Biocomput. pp 351–362

60. Tang J, Karhinen L, Xu T et al (2013) Target inhibition networks: Predicting selective combinations of druggable targets to block cancer survival pathways. PLoS Comput Biol 9:e1003226

61. Ranjan G, Zhang ZL (2013) Geometry of complex networks and topological centrality. Physica A 392:3833–3845

62. Estrada E, Hatano H (2010) Resistance distance, information centrality, node vulnerability and vibrations in complex networks. In: Estrada E, Fox M, Higham D, Oppo GL (eds) Network science: complexity in nature and technology. Springer, New York, pp 13–29

63. Yang W, Soares J, Greninger P et al (2013) Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 41(Database issue):D955–61

64. Cappuccio A, Zollinger R, Schenk M et al (2015) Combinatorial code governing cellular responses to complex stimuli. Nat Commun 6:e0127433

CRISPR/Cas-Mediated Knockin in Human Pluripotent Stem Cells

Nipun Verma, Zengrong Zhu, and Danwei Huangfu

Abstract

Fluorescent reporter and epitope-tagged human pluripotent stem cells (hPSCs) greatly facilitate studies on the pluripotency and differentiation characteristics of these cells. Unfortunately traditional procedures to generate such lines are hampered by a low targeting efficiency that necessitates a lengthy process of selection followed by the removal of the selection cassette. Here we describe a procedure to generate fluorescent reporter and epitope tagged hPSCs in an efficient one-step process using the CRISPR/Cas technology. Although the method described uses our recently developed iCRISPR platform, the protocols can be adapted for general use with CRISPR/Cas or other engineered nucleases. The transfection procedures described could also be used for additional applications, such as overexpression or lineage tracing studies.

Key words Human pluripotent stem cells (hPSCs), Gene targeting, CRISPR/Cas, Homologous recombination, Knockin, Fluorescent reporter, Epitope tag

1 Introduction

Human pluripotent stem cells (hPSCs), which include embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs), have two unique characteristics: the capacity for unlimited self-renewal in culture and the ability, called pluripotency, to form any cell type that is present in an adult human. Since the isolation of hESCs in 1998, there has been much progress in the development of protocols to differentiate hPSCs into specific cell types. As a result hPSCs have emerged as a valuable tool for studying human development and for modeling human disease. Furthermore there is great potential to use hPSCs to generate clinically important cell types, which can be transplanted into patients to replace those cells lost or damaged due to disease [1, 2].

The utility of hPSCs can be further enhanced through genetic modification of these lines. Through homologous recombination exogenous sequences can be inserted into the genome to generate a variety of "knockin" alleles, from a single nucleotide change to

introduce or correct a disease associated mutation, to the insertion of small (less than 100 base pairs, bp) epitope tags (e.g., hemagglutinin, 3× FLAG, V5 tags) or even the insertion of large (hundreds of bp) transgene constructs such as fluorescent reporters (e.g., GFP, mOrange) and antibiotic selection cassettes. The ability to insert epitope or fluorescent tags into the genome would greatly facilitate hPSC-based in vitro and in vivo (after transplantation in animal models) studies. The insertion of epitope tags into specific loci creates recombinant hybrids that contain a polypeptide affinity tag and enable efficient purification of target proteins. The development of these fusion proteins facilitates a variety of downstream experiments including: western blot, immunoprecipitation (IP), chromatin-immunoprecipitation (ChIP), and immunofluorescence staining, and is particularly useful when antibodies for the protein of interest are substandard or not available. Lineage-specific knockin fluorescent reporters are valuable tools because they allow real-time observation of gene expression dynamics, cell lineage tracing, and the isolation of a specific cell population for further analysis. hPSCs reporter lines have already been used to identify mediators of pluripotency [3] as well as characterize the culture conditions that stabilize different pluripotent states [4].

Creating knockin alleles in hPSCs used to be extremely challenging because of the low transfection efficiency [5] and low rate of spontaneous homologous recombination [6]. Thus traditional knockin strategies rely on the use of a drug-resistance cassette that allows for the enrichment of cells with the correct integration. However, due to potential interference between the drug-resistance cassette and nearby genes, the selection cassette must be removed prior to using these cells for experimental studies [7]. We have ourselves generated a knockin reporter allele at the *OCT4* locus using antibiotic selection [8]. Notably correctly targeted clones with the drug selection cassette did not express GFP. Only upon Cre-mediated deletion of the drug selection cassette, did all clonal lines show proper co-expression of pan-cellular eGFP with OCT4. Thus the low targeting efficiency of hPSCs necessitates an antibiotic selection cassette, but this in turn requires additional targeting, isolation, and characterization of clonal lines, effectively doubling the time and effort required to make the knockin line.

The recent emergence of programmable site-specific nucleases has greatly improved the efficiency of genome engineering in hPSCs. Zinc finger nucleases (ZFNs), transcription-activator like endonucleases (TALENs), and the CRISPR/Cas9 nuclease can all act as genomic scissors to produce a double strand break (DSB) at the desired genome loci [9–11]. Without a repair template, the DSB is repaired by the nonhomologous end joining (NHEJ) pathway that often generates small insertions or deletions (indels). These hPSC lines with site-specific knockout mutations can be used to determine the contribution of particular genes for development

and disease. In the presence of a repair template, DSB repair with error-free homology directed repair (HDR) can be used to generate knockin alleles through targeted integration of exogenous DNA sequences into the desired loci.

The CRISPR/Cas9 system was discovered as a form of adaptive immunity in bacteria, where it is used to destroy exogenous nucleic acids of invading viruses [12], and it has recently been adapted for genome engineering in mammalian cells. In this system the CRISPR chimeric guide RNA (gRNA) recognizes a 20-nucleotide (nt) DNA sequence upstream of the 5′-NGG-3′ protospacer adjacent motif (PAM) and directs the DNA endonuclease Cas9 for site-specific cleavage [13–16]. Cas9 produces a DNA DSB, which in the presence of a double stranded DNA (dsDNA) plasmid or single stranded DNA (ssDNA) template can promote HDR to efficiently incorporate exogenous sequences, such as a fluorescent reporter tag, into the specific genomic locus in hPSCs [8, 17–20]. We have developed an efficient genome-editing platform in hPSCs, called iCRISPR [21, 22]. Through TALEN-mediated gene targeting, a doxycycline inducible Cas9 cassette and the reverse tetracycline transactivator (M2rtTA) are integrated into both alleles of the endogenous *AAVS1* locus, and thus allow robust Cas9 expression in established clonal lines (referred to as iCas9 hPSCs) upon doxycycline treatment. With transfection of appropriate gRNAs we are able to generate site-specific knockout mutations. In addition, with co-transfection of repair templates, iCRISPR has also allowed us to specifically alter the sequence of a small number of nucleotides, and to insert a knockin epitope tag or a large transgenes such as a fluorescent tag. Most importantly due to the increased efficiency of the iCRISPR system it is not necessary to use selection to enrich for correctly integrated alleles. Thus it is feasible to generate knockin alleles in an efficient one-step procedure [8].

In this methods chapter we describe the procedure for using our iCRISPR system to generate knockin alleles in hPSCs. We specifically describe the procedure for generating epitope-tagged and fluorescent-tagged hPSC lines; however, the same knockin strategy could be used for a variety of other inserts. Furthermore, we believe the transfection approach that we use to generate the lines could also be used for other purposes, for example to deliver mRNA or plasmids for overexpression or lineage tracing studies.

2 Materials

The following materials are needed.

1. hPSC medium: DMEM/F12 medium (Life Technologies), 20% KnockOut Serum Replacement (Life Technologies), 1× Non-Essential Amino Acids (Life Technologies), 1× GlutaMAX

(Life Technologies), 100 U/mL Penicillin/100 μg/mL Streptomycin (Life Technologies), 0.055 mM 2-mercaptoethanol (Life Technologies), 10 ng/mL recombinant human basic FGF (Life Technologies).

2. TrypLE Select enzyme (Life Technologies).

3. Fetal Bovine Serum (FBS, Sigma-Aldrich).

4. Dimethyl Sulfoxide (DMSO, Santa Cruz Biotechnology).

5. Rho-associated protein kinase (ROCK) inhibitor Y-27632 dihydrochloride (Selleck Chemicals).

6. Genomic DNA lysis buffer: 10 mM Tris pH 8, 0.45 % NP40, 0.45 % Tween 20, and 100 μg/mL Proteinase K.

7. Lipofectamine RNAiMAX (Life Technologies).

8. Lipofectamine 3000 (Life Technologies).

9. Herculase II Fusion DNA Polymerase (Agilent Technologies).

10. MEGAshortscript T7 Transcription kit (Life Technologies).

11. MEGAclear Transcription Clean-Up Kit (Life Technologies).

12. DNeasy Blood and Tissue Kit (Qiagen).

13. Ethidium Bromide (EtBr).

14. Agarose (Invitrogen).

15. Purelink Gel Purification Kit (Life Technologies).

16. Pierce Crosslink Magnetic IP/Co-IP kit (Pierce Antibodies).

17. FLAG antibody, clone M2 (Sigma-Aldrich).

18. TBST: 20 mM Tris–HCl (pH 7.6), 137 mM NaCl, 0.1 % Tween 20.

19. Blocking Buffer: 5 % nonfat milk in TBST.

20. NuPAGE LDS Sample Buffer 4× (Life Technologies).

21. NuPAGE® Sample Reducing Agent 10× (Life Technologies).

22. NuPAGE Novex 3–8 % Tris–Acetate Protein Gels (Life Technologies).

23. NuPAGE Tris–Acetate SDS Running Buffer (20×) (Life Technologies).

24. Nitrocellulose Pre-Cut Blotting Membranes (Life Technologies).

25. Amersham ECL Prime Western Blotting Detection Reagent (Amersham).

26. Amersham Hyperfilm (Amersham).

27. Amersham Hypercassette (Fisher).

3   Methods

Here we describe the procedures to generate epitope and reporter tagged hPSC lines using the iCRSIPR platform. In the iCRISPR system a doxycycline inducible Cas9 cassette has been targeted into the endogenous *AAVS1* locus. Upon doxycycline treatment, Cas9 is expressed in all cells. After transfection of the site-specific gRNAs and the donor HDR template with the exogenous DNA sequence flanked by the homology arms, a DSB is created and HDR leads to incorporation of the insert into the genome. For single nucleotide alternations, a ssDNA template with homology arms of ~40–80 nt is generally used. Due to the small size of the epitope tags (less than 100 nt), the exogenous sequence can also be provided as a ssDNA, whereas the larger reporter tag (hundreds of bp) can be cloned easily into a donor plasmid for delivery into the cells. In both cases selection is not required and knockin alleles can be efficiently generated in a single step (Fig. 1). We have used the following protocols for HUES8 (NIHhESC-09-0021) and MEL-1 (NIHhESC-11-0139) hESC lines. It may be necessary to adjust the protocols for different hPSC lines or for hPSCs cultured under different conditions such as feeder-free TeSR and Essential 8 (E8) culture conditions [23, 24].
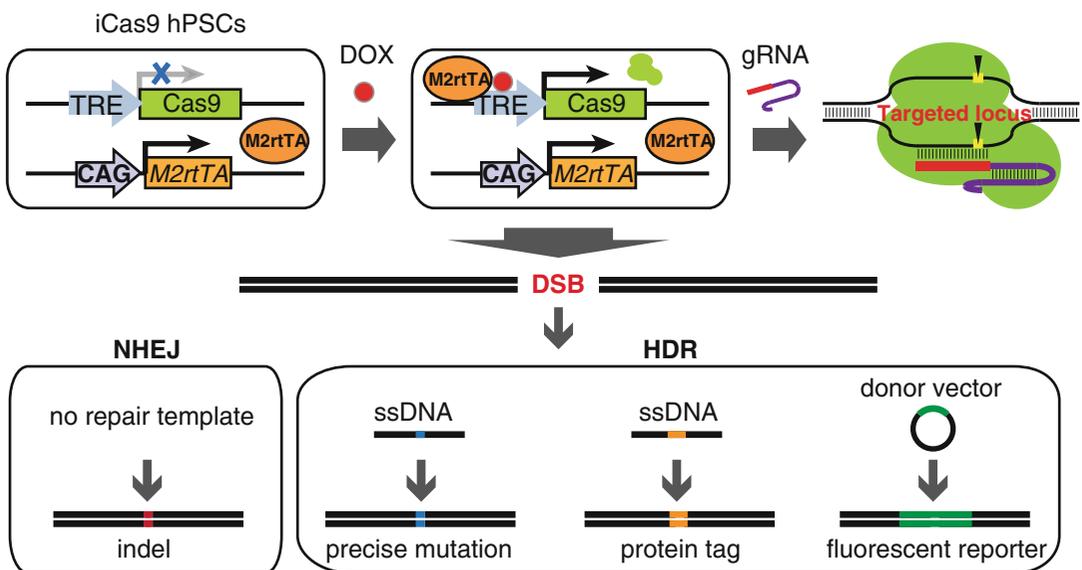
Fig. 1 The iCRISPR platform for genome editing in hPSCs. Doxycycline treatment induces Cas9 expression in iCas9 hPSCs. After transfection of gRNA, Cas9 is guided to the target locus via Watson-Crick base pairing and induces DNA DSBs. In the absence of a repair template, repair of DSBs by NHEJ often results in indels that can be used to knockout a target gene in hPSCs. Alternatively, in the presence of repair templates, either short ssDNA or long dsDNA donor vectors, HDR can be employed to incorporate exogenous sequences into the endogenous loci to introduce precise nucleotide mutations, protein tags, or fluorescent reporters into the target loci

3.1 hPSC Culture	1. hPSCs are cultured on an irradiated mouse embryonic fibroblast (iMEF) feeder layer in hPSC media.

2. iMEFs are plated 1 day before seeding hPSCs. Tissue culture dishes are coated with 0.1% gelatin for 30 min at 37 °C prior to plating the iMEF feeder layer.

3. hPSCs cultures are generally passaged at 1:6 or 1:12 split ratios every 4–6 days. To passage hPSCs, colonies are disaggregated by treating with TrypLE Select enzyme for 5 min at 37 °C.

4. Dissociated hPSCs are collected in hPSC media and spun down at $200 \times g$ for 5 min. After spinning hPSCs are resuspended in fresh hPSC media and seeded on iMEF coated plates. 5 μM ROCK inhibitor Y-27632 is added into the culture medium when passaging or thawing frozen cells.

5. Cell lines are frozen down in freezing media consisting of 40% FBS, 10% DMSO and 50% hPSC media.

3.2 Generation of Knockin Reporter hPSC Lines Using iCRISPR

The generation of knockin reporter hPSC lines using the iCRISPR system involves first the induction of Cas9 expression through administration of doxycycline followed by the co-transfection of the gRNA and donor vector into the iCas9 cells. The transfected gRNA forms a complex with the Cas9 protein and directs it to the target locus, where it creates a DSB. HDR can then be employed to efficiently incorporate the exogenous reporter sequence into the target locus. After dissociating and re-plating, single-cell colonies are isolated, expanded and screened through PCR for correct reporter integration. Correctly targeted clonal lines are further validated and differentiated to assess faithful reporter activity (Fig. 2).
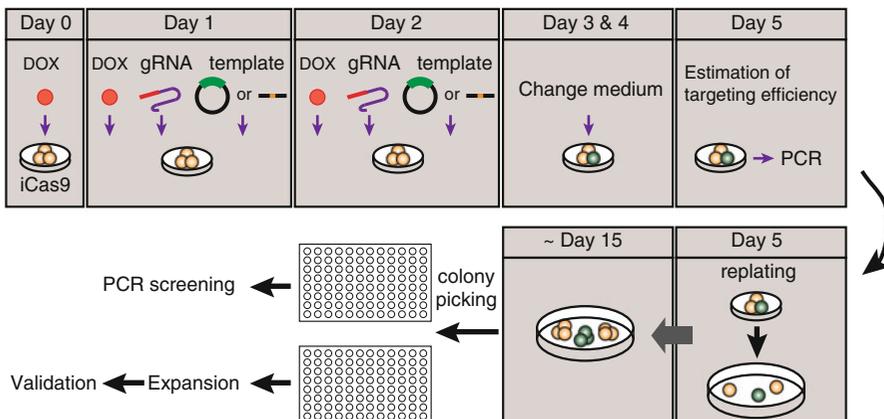
Fig. 2 The workflow for iCRISPR-mediated knockin in hPSCs. Doxycycline treatment 24 h before the first transfection induces Cas9 expression in iCas9 hPSCs at Day 0. Co-transfection of gRNA and repair templates, either donor vectors or ssDNAs, on Day 1 and Day 2 results in the incorporation of exogenous sequences into the endogenous loci. After PCR genotyping on Day 5, cells with the highest targeting efficiency will be re-plated as single-cells at a low density (~2000–5000 cells /100 mm dish) and allowed to grow for ~10–15 days. On Day 15, single-cell colonies will be picked, amplified, passaged into duplicated wells and subject to PCR screening. Correctly targeted clonal lines will be expanded and further validated

3.2.1 Donor Vector Design and Construction

Traditionally, the donor plasmid contains two homology arms flanking a promoterless reporter gene and a drug-resistance cassette driven by a constitutively active promoter. The two arms contain sequences that are homologous to the target locus. The constitutively active drug-resistance cassette enables enrichment of correctly targeted clones. However to avoid interference between the drug resistance cassette and neighboring genes, the drug-resistance cassette needs to be removed after identification of correctly targeted clones. Due to the high efficiency of the iCRISPR system, knockin reporter lines can be generated without the use of a drug selection cassette. This significantly reduces the time and effort required for establishment of knockin reporter hPSC lines.

1. Choose the target site for reporter integration in a way that ensures faithful reporter expression as well as minimizes the potential impact on endogenous gene expression and/or protein function. There are two general strategies for reporter gene integration, protein fusion (also called translational fusion) and promoter fusion (also called transcriptional fusion) (Fig. 3). For protein fusion, a reporter gene is integrated into the same reading frame of the target gene, transcribed as a single mRNA driven by the endogenous promoter and expressed as a single fusion protein. The protein fusion reporter is useful for monitoring protein subcellular localization, protein dynamics, and protein–protein interactions of the target gene. However, the fused reporter may affect the folding, stability, and function of the endogenous protein. Without previous knowledge about the functional domain and protein structure of the target gene, we recommend inserting the reporter gene either at the N-terminus immediately after the start codon or at the C-terminus immediately after the coding sequence. For promoter fusion, a reporter gene is also integrated into the same reading frame and transcribed together with the target gene as a single mRNA driven by the endogenous promoter. However, different from protein fusion reporter, promoter fusion reporter is preceded by either an IRES (internal ribosome entry site) or a 2A peptide sequence (Fig. 3). Consequently, the reporter and target genes will be translated into two separate peptides. Promoter fusion reporter reflects the endogenous gene expression pattern and usually has minimal effects on the endogenous gene expression or protein function. However, it is not suitable to monitor the subcellular localization, protein dynamics, and protein–protein interactions of the target gene. Whether to make a protein fusion or a promoter fusion reporter depends on the intended application. For monitoring gene expression during hPSC differentiation, we are primarily concerned with dynamic changes in gene expression and less about the subcellular localization of the protein. Thus for this purpose we generally use the promoter fusion strategy and integrate the reporter gene at
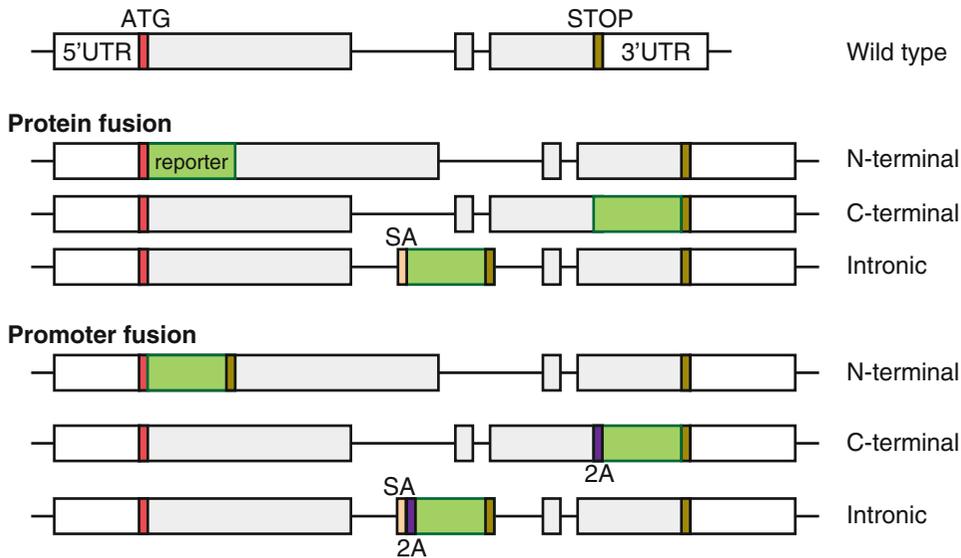
Fig. 3 Schematics of protein and promoter fusion knockin strategies. Here we illustrate typical targeting strategies for creating protein and promoter fusion knockin alleles. For a protein fusion reporter, the reporter gene (in *green*) is fused in-frame to the N-terminus of the endogenous gene immediately after the start codon (ATG, in *red*), to the C-terminus immediately after the last coding sequence or in the intronic region using a splicing acceptor sequence (SA, in *orange*). For a promoter fusion reporter, the reporter gene and the endogenous locus is linked through a 2A peptide (2A, in *purple*) and thus two separate peptides will be produced. The promoter fusion reporter can also be integrated into the N-terminus, C-terminus or intronic region of the endogenous gene. Boxes are exons, with open boxes indicating the untranslated region (UTR), filled grey boxes indicating the coding sequence (CDS); connecting lines are introns

the C-terminus immediately after the coding sequence of the endogenous gene as this will have the least effect on endogenous protein function (Fig. 4a).

2. After deciding the integration locus, PCR-amplify the two homologous arms (HA-L and HA-R) flanking the integration locus using high fidelity Herculase II Fusion DNA Polymerase (*see* **Note 1**) and clone into a backbone vector. We generally use pBluescript, which contains multiple cloning sites and is relatively small, as the backbone vector.

3. While the optimal length of the homologous arms has not been thoroughly investigated, approximately 500–1000 bp of homology arms on each side of the reporter gene are recommended (*see* **Note 2**).

4. Insert the reporter, for example a fluorescent protein reporter, preceded by a 2A self-cleavage sequence (2A-reporter) between the two homology arms. Make sure the 2A-reporter sequence is fused in-frame to the last coding sequence of the target gene (Fig. 4b).
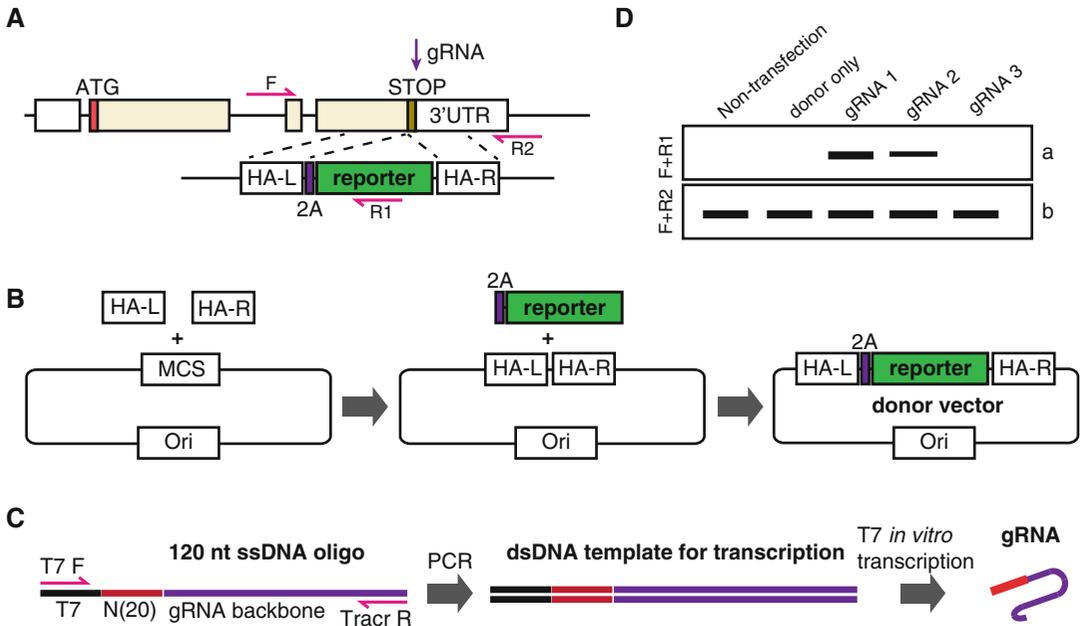
Fig. 4 Generation of knockin reporter hPSC lines using iCRISPR. (**a**) Schematics of the targeting strategy. In the presence of the donor vector, HDR results in the replacement of stop codon with 2A-reporter sequence. The PCR primers (F + R1 and F + R2) used for genotyping are indicated with *red arrows*. (HA-L and HA-R indicate left and right homology arms). (**b**) Schematics of the cloning strategy. The left and right homology arms (HA-L and HA-R) and the 2A-reporter gene are cloned sequentially into the backbone vector. (*MCS* multiple cloning site). (**c**) Schematics of the gRNA production. A 120 nt ssDNA containing a 20 nt T7 promoter sequence (T7, in *black*), a variable 20 nt gRNA recognition sequence ((N)20, in red) and a constant gRNA backbone sequence (in *purple*) is PCR amplified using T7 F and Tracr R universal primers (*red arrows*). The dsDNA PCR products will then be used as templates for in vitro transcription to produce the gRNAs. (**d**) Estimation of targeting efficiency. The PCR primers (F + R1 and F + R2) are indicated in panel (**a**). Comparison of the PCR product (band a) and (band b) allows an approximate estimation (a/(a + b) x 100 %) of the targeting efficiency by each gRNA

3.2.2 Design and Production of gRNAs

We routinely use the CRISPR design tool developed by the Feng Zhang group at MIT (http://crispr.mit.edu/). The software not only identifies all possible CRISPR targets in an input DNA sequence but also uncovers potential off-target sites, thus predicting gRNAs with the highest targeting specificity. For each target locus, we recommend designing 3 gRNAs and selecting the gRNA that shows the highest targeting efficiency for establishment of clonal lines (estimation of the targeting efficiency is described in **step 5** of Subheading 3.2.3).

We have found that it is faster and more cost effective to generate gRNAs using an oligonucleotide, than through cloning into the commonly used gRNA-expressing plasmids. First, a 120 nt ssDNA oligo, which includes a T7 promoter sequence, a variable 20 nt gRNA recognition sequence (N)20 (in red) and a constant gRNA backbone sequence can be directly synthesized and bought from commercial vendors. Next PCR-amplify the oligonucleotide

using the T7 F and Tracr R universal primers and use the PCR products as templates for in vitro transcription to produce the gRNAs (Fig. 4c).

1. Use the online CRISPR design tool (http://crispr.mit.edu/) to design 3 gRNAs that target sequences at the appropriate genomic site. It is highly recommended to choose gRNAs that do not recognize any sequence present in the donor vector to prevent undesired CRISPR/Cas9-mediated mutagenesis after correct reporter gene integration.

2. Order the 120 nt ssDNA oligos containing the desired 20 nt gRNA sequences directly from commercial vendors. Desalted purification is sufficient.

3. PCR-amplify the 120 nt ssDNA oligos using the T7 F and Tracr R universal primers using Herculase II Fusion DNA Polymerase to produce dsDNA templates for T7 in vitro gRNA transcription.

4. Examine the PCR product through electrophoresis using 2.5 % Agarose gel. Gel purify the desired dsDNA templates if non-specific PCR products are detected.

Primer	Sequence
T7 F	TAATACGACTCACTATAGGG
Tracr R	AAAAGCACCGACTCGGTGCC

PCR Reaction mix (50 μL)

Component	Amount (μL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
T7 F (10 μM)	1.25
Tracr R (10 μM)	1.25
120-nt ssDNA oligo (25 nM)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2–31	94 °C, 20 s	60 °C, 20 s	72 °C, 1 min
32			72 °C, 2 min

5. Use the MEGAshortscript T7 Transcription kit for gRNA synthesis. In vitro *t*ranscription mix (20 µL)

Component	Amount (µL)
T7 ATP	2
T7 CTP	2
T7 GTP	2
T7 UTP	2
T7 10× buffer	2
T7 enzyme mix	2
dsDNA template	8

Incubate for 4 h to overnight at 37 °C

Add 1 µL TURBO DNase and incubate for 15 min at 37 °C to digest the dsDNA template.

6. Purify gRNA using the MEGAclear Transcription Clean-Up Kit following the manufacturer's instructions, and elute gRNAs (typically ~50–100 µg) in 100 µL RNase-free water. When possible adjust concentration to 320 ng/µL (10 µM) and store at –80 °C until use.

3.2.3 Co-transfection of gRNA and Donor Vector in iCas9 hPSC Cells

Efficient co-transfection of gRNA and dsDNA donor vector into the iCas9 hPSC cells is very important for the generation of knockin reporter hPSC lines using iCRISPR system. We have tested several commonly used transfection reagents along with electroporation and find that the highest and most consistent co-transfection efficiency (~20%) is achieved using Lipofectamine 3000 (*see* **Note 3**).

1. *Day 0.* Plate iMEFs on a gelatin-coated 24-well plate and treat 60% confluent iCas9 hPSC with hPSC medium containing 2 µg/mL doxycycline. This will allow optimal Cas9 expression at the time of transfection on Day 1.

2. *Day 1.* Wash the iCas9 cells with PBS w/o Ca^{2+} and Mg^{2+} and then dissociate hPSCs with TrypLE for 5 min at 37 °C. Once the cells are disaggregated, collect the cells in 10 mL of hPSC media and spin down at $200 \times g$ for 5 min. Resuspend cells at ~$0.25–1 \times 10^6$ cells/mL in hPSC medium with 5 µM ROCK

inhibitor and 2 μg/mL doxycycline (*see* **Note 4**). Re-plate 0.5 mL of cell suspension into each well of a 24-well plate. Prepare duplicated wells for each gRNA used, and wells for the non-transfection control and the donor-only control. One well of the duplicated wells will be used for estimation of the targeting efficiency and the other well will be used for re-plating.

First transfection. For each gRNA, prepare separately:

Mix 1: 50 μL Opti-MEM + 1 μL gRNA (320 ng) + 2 μL P3000 + 2 μL donor plasmid (2.5 μg).

Mix 2: 50 μL Opti-MEM + 3 μL Lipo 3000 reagent.

Mix 1 + 2, incubate for 5 min at RT and add 50 μL mixture dropwise into 0.5 mL hPSCs dilution in both duplicate wells.

For the non-transfection control, omit both gRNA and donor template.

For the donor-only control, add only donor template.

3. *Day 2.* Optional. Change hPSC medium with 2 μg/mL doxycycline and perform the 2nd transfection using the same procedure as the 1st transfection.

4. *Day 3–4.* Change hPSC medium daily.

5. *Day 5.* Estimation of targeting efficiency.

For lineage-specific genes that are only expressed in differentiated cell types, the targeting efficiency can be roughly estimated by PCR amplification of the targeted alleles. Collect genomic DNA from one of the duplicated wells using the DNeasy Blood & Tissue Kit. PCR-amplify the targeted alleles using an external primer (F) and an internal primer (R1), and the non-targeted allele using two external primers (F and R2) (Fig. 4d). Although primers F + R2 may also amplify the targeted allele, in our experience the non-targeted allele is preferentially amplified due to its smaller size. Identify the gRNA with the highest targeting efficiency and use the corresponding well for expansion and screening of clonal lines.

PCR Reaction mix (50 μL)

Component	Amount (μL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
Forward primer (10 μM)	1.25
Reverse primer (10 μM)	1.25
Genomic DNA (100 ng/μl)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2–36	94 °C, 20 s	N[a] °C, 20 s	72 °C, 30 s
37			72 °C, 2 min

[a]Annealing temperature of the primer pair, ideally between 55 °C and 65 °C

For genes that are expressed in undifferentiated hPSCs, the percentage of cells expressing the fluorescence reporter tag can be directly assessed through flow cytometric analysis.

3.2.4 Expansion and Screening of Clonal lines

1. *Re-plating.* hPSCs identified with the highest targeting efficiency in Subheading 3.2.3 are dissociated into single cells using TrypLE, resuspended in hPSC medium with 5 μM ROCK inhibitor and re-plated at ~2000–5000 cells per 100 mm dish into dishes pre-seeded with iMEF feeders.
For genes that are expressed in undifferentiated hPSCs, enrich correctly targeted cells through Flow Activated Cell Sorting (FACS) and then re-plate them at a density of 10,000 cells per 100 mm dish.

2. Do not change medium the next day.

3. Starting from 2 days after re-plating, change hPSC medium every day until colonies growing from single cells reach ~2 mm in diameter (~10 days).

4. *Colony picking.* Remove hPSC medium and add 10 mL PBS w/o Ca²⁺ and Mg²⁺ into the dish for a clearer visibility of the colonies. Mechanically dissociate the hPSC colonies into small clusters using 200 μL pipette tips, pick and plate single colonies in an uncoated 96-well plate containing 100 μL of hPSC medium with 5 μM ROCK inhibitor. Each clone is further disaggregated by pipetting up and down 5 times and re-plated in duplicated 96-well plates (50 μL each) pre-seeded with iMEFs and containing 100 μL of hPSC medium with 5 μM ROCK inhibitor. Alternatively, colonies could first be cultured in one 96-well plate and then split into two 96-well plates when confluent (in ~5 days).

    Depending on the estimated targeting efficiency in Subheading 3.2.3, pick between 96 (targeting efficiency >5 %) to 384 clones (targeting efficiency <5 %). Media change and passaging can be performed using a multichannel aspirator and a multichannel pipette.

    Change hPSC media for the 96-well plates daily. Use one of the duplicated 96-well plates for PCR screening in 3 days and the other plate for maintenance.

5. *Genomic DNA extraction.* Remove media from the wells, wash once with PBS w/o Ca²⁺ and Mg²⁺ and add 35 µL of genomic DNA lysis buffer. After 5 min at RT, transfer the cell lysates into a 96-well PCR plate. Seal the plate using a PCR film sticker and incubate at 55 °C for 2 h followed by 5 min at 95 °C to inactivate Proteinase K. Keep samples at 4 °C before the PCR screening step (*see* **Note 5**).

6. *PCR screening.* PCR-amplify using the same external primer (F) and internal primer (R1) used in Subheading 3.2.3. Run 25 µL of the PCR product in a 1% agarose gel stained with EtBr. PCR products of the right size suggest proper targeting in clonal lines.

7. *Expansion of clonal lines.* Amplify the desired clonal lines from the maintenance plate for further validation.

3.2.5 Validation of Established Clonal Lines

After targeting, isolation of clonal lines and PCR genotyping, correctly targeted clonal lines are expanded and further validated. We recommend immunohistochemistry for expression of pluripotency markers (e.g., OCT4, SOX2, and NANOG), teratoma assay for a functional assessment of pluripotency, karyotype testing, and differentiation into specific lineages to confirm faithful reporter activity. We highly recommend performing Sanger sequencing of both the targeted and the non-targeted allele at the target locus. We generally detect the expected sequence at the targeted allele. However, indel mutations are sometimes detected in the non-targeted allele, and clones with these undesired mutations should be discarded.

Due to mismatch tolerance of CRISPR gRNA paring [18], there are concerns about the potential off-target mutagenic effects of the CRISPR/Cas9 system. The online CRISPR design tool (http://crispr.mit.edu/) can be used to identify the potential off-targets falling in coding sequences for each gRNA and we typically select 10 top candidates for further analysis. PCR-amplify the genomic region (~500 bp) flanking the CRISPR off-target site using Herculase II Fusion DNA Polymerase and analyze by Sanger sequencing using a primer that binds within the PCR product. It is worth noting that our analysis so far has not revealed any off-target mutations at sites without perfect complementarity to the CRISPR/Cas9 target sequence. In agreement with our findings, a recent high-coverage whole-genome sequencing study failed to detect significant incidence of off-target mutations in CRISPR-targeted hPSC lines [25].

3.3   Generation of Knockin Epitope-tagged hPSCs Using iCRISPR

The procedure to generate knockin epitope tagged hPSCs is similar to the procedure used to create reporter tags (Subheading 3.2). Expression of Cas9 is induced in iCas9 cells following administration of doxycycline. A gRNA for the target site and the exogenous sequence are delivered into the cells. Site-specific DSBs produced by Cas9 promote HDR and integration of the epitope tag into the

genome. After two transfections the cells are re-plated at clonal densities. Clonal lines are then isolated, expanded and genotyped using PCR.

One key difference from Subheading 3.2 is that due to the smaller size of the epitope sequence, a ssDNA can be used to provide the HDR template (the epitope tag flanked by the homology arms). In addition a different transfection reagent, Lipofectamine RNAiMax is used to deliver the ssDNA and gRNA into the cells. Finally clonal lines identified by PCR to have correct integration, are further validated using Sanger sequencing, IP and western blot.

3.3.1 Design of CRISPR gRNAs and Donor ssDNA

For the generation of tagged proteins the location of the tag in reference to the protein is very important. The tag should be located so as to have minimal effect on tertiary structure and biological activity of the protein. At the same time the tag should be readily accessible on the surface of the natively folded protein. In deciding where to insert the tag it is best to follow previous studies that have already validated the generation of a functional fusion protein. In the absence of a previous study, the safest strategy would be to generate lines that have the tag epitope attached to the N-terminal of the target protein, as well as lines that have the tag epitope attached to the C-terminal of the target protein (Fig. 5a).
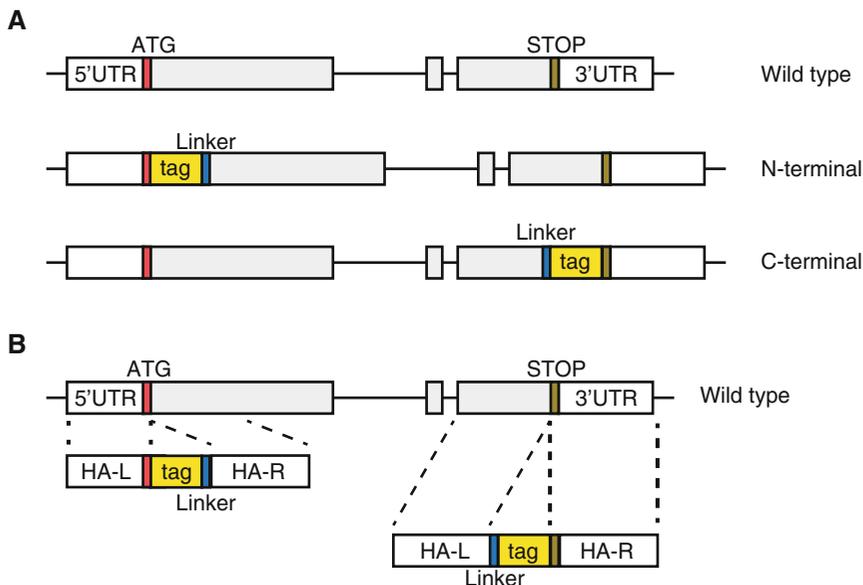
Fig. 5 Generation of epitope tag knockin hPSC lines using iCRISPR. (**a**) Knockin epitope tags can be targeted to the N-terminal or C-terminal of a protein. At the N-terminal the epitope tag is inserted right after the start codon and separated from the endogenous sequence by a linker sequence. At the C-terminal the epitope tag is inserted immediately before the stop codon with a linker sequence separating the epitope tag from the endogenous sequence. (**b**) Schema of the ssDNA donors used to generate knockin epitope tags

After these lines have been generated one can compare them to see if the tag influences the stability or biological function of the protein. This could be done using western blot, co-IP with known interacting proteins, and for DNA-binding proteins, ChIP-qPCR.

When targeting the N-terminal of the protein, 3-4 gRNAs are chosen to cut near the start codon, so that after integration the start codon will immediately precede the epitope tag. When targeting the C-terminal of a protein, 3–4 gRNAs are designed to cut near the stop codon of the protein, so that after integration the epitope tag will be immediately before the stop codon (Fig. 5a). Please refer to Subheading 3.2.2 for detailed instructions on the design and production of CRISPR gRNAs.

3.3.2 Design of Donor ssDNA

The donor template consists of three components: the left and right homology arms, the epitope tag sequence, and a short linker sequence. The smaller size of the epitope tag compared to a reporter transgene, allows smaller homology arms (less than 100 nt) to be used for the targeting. Thus the overall size of the donor template is typically between 150 and 300 nt, and a ssDNA donor, instead of a plasmid, can be used for the targeting experiment.

1. Each of the homology arms (left and right) is around 65 nt in length (*see* **Note 6**).

2. For targeting the N-terminus, the left homology arm should stop after the start codon of the protein, and the right homology arm should include the DNA sequence immediately after the start codon (Fig. 5b). The sequence of the epitope tag is inserted between the homology arms (*see* **Note 7**), followed by a short linker sequence (e.g., 12 nt long). The linker sequence separates the epitope tag from the endogenous protein. This prevents the epitope tag from interfering with proper folding of the endogenous protein [26], and increases the accessibility of the epitope to antibodies [27]. A linker sequence we have used is: GGAGGACTCGAC, which produces a four amino acid sequence of Gly-Gly-Leu-Asp.

3. If targeting the C-terminus, the left arm should contain the sequence up to, but not including, the stop codon. This is followed by the linker sequence and then the epitope tag. Lastly the right homology arm should begin with the stop codon (Fig. 5b).

4. The final ssDNA oligo, containing the two homology arms, a linker sequence and the epitope tag, can be ordered directly from commercial vendors. Desalted purification is sufficient.

3.3.3 Co-transfection of Donor ssDNA and CRISPR gRNA

The procedure for passaging the cells for co-transfection is the same as described in Subheading 3.2.3. However, the transfection mixture is different. Perform co-transfection of donor ssDNA and CRISPR gRNA as follows:

Mix 1: 50 µL Opti-MEM + 1 µL gRNA (320 ng) + 5 µL ssDNA (1.5 µg).

Mix 2: 50 µL Opti-MEM + 3 µL Lipofectamine RNAiMAX.

Mix 1 + 2, incubate for 5 min at RT and add 50 µL mixture dropwise into 0.5 mL dissociated hPSCs in one well of a 24-well plate.

The same procedure presented in **step 5** of Subheading 3.2.3 can be used to evaluate the targeting efficiency for the epitope-tag.

3.3.4 Expansion and PCR Based-screening of Clonal Lines

The procedure for isolating clonal lines, PCR screening and expansion of clonal lines is the same as described in Subheading 3.2.4.

3.3.5 Sequencing Verification of Targeted Clones

We highly recommend examining both the targeted and non-targeted allele at the target locus by Sanger sequencing. This enables us to verify that the epitope tag sequence is intact and properly integrated in-frame with the protein sequence. We also verify that the non-targeted allele does not contain any unwanted mutations.

1. Collect genomic DNA of correctly targeted lines using the DNeasy Blood and Tissue kit.

2. PCR-amplify the targeted and non-targeted alleles using a pair of external primers.

PCR Reaction mix (50 µL)

Component	Amount (µL)
ddH$_2$O	35.5
5× Herculase II reaction buffer	10
dNTP mix (25 mM)	0.5
Forward primer (10 µM)	1.25
Reverse primer (10 µM)	1.25
Genomic DNA (100 ng/µl)	1
Herculase II fusion DNA polymerase	0.5

PCR cycling conditions

Cycle number	Denature	Anneal	Extend
1	94 °C, 2 min		
2-36	94 °C, 20 s	N[a] °C, 20 s	72 °C, 30 s
37			72 °C, 2 min

[a]Annealing temperature of the primer pair, ideally between 55 °C and 65 °C

3. Run 25 μL of the PCR reaction on a 1% agarose gel stained with EtBr.

4. Visualize the gel using a UV lamp. For correctly targeted clones there should be two bands, representing the targeted allele and the non-targeted allele. The targeted allele should be larger than the non-targeted allele by the size of the epitope tag (generally less than 100 bp).

5. Cut out the individual bands and purify DNA using the Purelink Quick Gel Extraction kit.

6. Use one of the PCR primers for Sanger sequencing, or use a new sequencing primer that is within the PCR product (*see* **Note 8**).

3.3.6 Validation of Targeted Clones Using Immunoprecipitation and Western Blot

The final, and most important, test to validate the epitope tagged lines is to perform IP for the tag and then verify that the target protein is also pulled down. This would involve first performing an IP using an antibody against the epitope tag followed by a western blot with the antibody to the protein of interest. If it is challenging to detect the target protein due to lack of quality antibodies, one may perform the western blot against the tag epitope to verify that the epitope tagged protein is the same size as the target protein. Below we provide a protocol routinely used in our laboratory for detection of 3×FLAG tagged proteins. We strongly encourage users to optimize the protocol for their protein of interest, as it may require different procedures for cell lysis, IP and western blot.

1. Expand the clonal lines into a 100 mm dish.

2. Use the Pierce Crosslink Magnetic IP/Co-IP kit for lysis and IP and follow the manufacturer's protocol. Lyse one 100 mm dish per IP with 600 μL of the provided lysis IP buffer. Incubate the cell lysate with the FLAG antibody for 16 h at 4 °C. And elute with 100 μL of the provided elution buffer.

3. For the western blot 30 μL total volume is run: 19.5 μL of IP sample, 7.5 μL LDS sample buffer and 3 μL of reducing agent. The samples should be denatured for 10 min at 70 °C prior to loading.

4. Samples are run on NuPAGE Novex 3–8% Tris–Acetate Protein Gels in NuPAGE Tris–Acetate SDS Running Buffer (20×) for 1 h at 120 V.

5. Transfer is performed overnight at 30 V onto a nitrocellulose membrane.

6. Wash the membrane in TBST and then incubate in blocking buffer (5% nonfat milk in TBST) for 30 min at room temperature.

7. Incubate the membrane with the FLAG antibody at a 1:2000 dilution overnight at 4 °C.

8. Wash the membrane 3 times for 5 min each with TBST, and then incubate with HRP conjugated anti-mouse antibody (1:5000) for 1 h at room temperature.

9. Finally develop the membrane using ECL reagent. Prior to exposure, excess liquid should be removed without drying the membrane. Transfer the membrane to a clean sealed bag. Expose onto an autoradiography film in a hypercassette.

4   Notes

1. Undesired mutations in homology arms arising during PCR amplification may decrease the targeting efficiency. We recommend using the high fidelity Herculase II Fusion DNA Polymerase for PCR amplification of homology arms and also later for PCR amplification of the gRNA in vitro transcription template. One may substitute Herculase with other commercially available high fidelity DNA polymerases.

2. Traditional targeting procedures that do not utilize site-specific nucleases require homology arms of at least 5–12 kb for the insertion of reporter and selection cassettes and the targeting efficiency can be further increased by using substantially longer homology regions (up to 100 kb) [28]. In most traditional targeting experiments, the donor plasmid typically contains a long (~10 kb) and a short homology arm (~2–4 kb) for efficient homologous recombination and also for convenient identification of correctly targeted alleles by PCR genotyping at the shorter homology arm side. With the use of engineered DNA endonucleases, repair templates with much shorter homology arms (~500–1000 bp on each side) have been used. We have found that approximately 500 bp combined length of the homology arms (~250 bp on each side) is necessary to support efficient CRISPR-mediated targeting of a 750 bp exogenous sequence into an endogenous locus.

3. In addition to Lipofectamine 3000 and RNAiMAX, a number of other transfection methods have also been shown to work efficiently in hPSCs [17, 19].

4. Based on our experience with HUES8 and MEL1 hESCs, we recommend testing several cell densities for the co-transfection of gRNA and donor vector, for example plating 100, 250, and 500 K cells per well in a 24-well plate and choosing the best condition for the establishment of clonal lines. After transfection, we regularly use a replating cell density of 2000–5000 cells per 100 mm dish. These conditions may need to be adjusted for other hPSC lines.

5. This simplified protocol enables the extraction of genomic DNA for PCR screening without the use of phenol/chloroform

extraction and purification. Combined with the use of multichannel pipettes, one can quickly screen a large number of clones. Based on our experience it is feasible for a single person to process up to 384 samples (four 96-well plates) at a time. This procedure works well as long as the primers are pre-tested before the targeting experiment.

6. We have used homology arms that are around 65 nt in length. The targeting efficiency may be increased by using longer homology arms, but this has not been systematically tested.

7. Thus far we have tested attaching a 3×FLAG epitope tag to the protein of interest. However this strategy can be used for attaching a variety of other epitope tags such as poly-Arg, poly-His, Strep-tag, c-myc-tag, HAT-tag, and His-tag. As long as the sequence of the tag remains under 100 bp the same procedure as presented above can be used. For larger tags the size of the homology arms and the delivery of the exogenous sequence (as a ssDNA or as a donor plasmid) may need to be adjusted. The specific protein tag to use will depend on the protein of interest as well as the intended downstream analysis.

8. If a nonsynonymous mutation is detected in the coding sequence of the protein in either the targeted or non-targeted allele we would advise to discard that particular clone. In some cases a mutation may occur in a non-coding regulatory region (e.g., promoter), in which case the clone may still be used if the mutation does not alter the expression of the affected allele (as determined by qRT-PCR for example).

5   Future Applications

Here we present the protocols to integrate epitope and reporter tags within the genome of hPSCs using the iCRISPR genome editing platform. The same knockin strategy could be used for additional purposes. For example, targeting a fluorescent or antibiotic tag into the coding region of a gene could be used to generate genetic knockouts. Specific deletions could be engineered to mimic a genetic disease or to dissect the contribution of particular protein domains for protein function. By inserting particular sequences one may even precisely modulate the DNA-binding, protein interaction or other specific activities of endogenous proteins. Furthermore, the gene targeting strategies presented here are also relevant to labs that use engineered nucleases other than CRISPR/Cas or employ different strategies to deliver these nucleases into hPSCs. Finally we provide robust protocols to achieve relatively high rates of delivery for ssDNA, dsDNA plasmids and gRNAs. These transfection procedures may have wider applicability such as potential uses in overexpression and lineage tracing studies.

Acknowledgements

References

1. Tabar V, Studer L (2014) Pluripotent stem cells in regenerative medicine: challenges and recent progress. Nat Rev Genet 15:82–92

2. Zhu Z, Huangfu D (2013) Human pluripotent stem cells: an emerging model in developmental biology. Development 140:705–717

3. Chia NY, Chan YS, Feng B et al (2010) A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature 468:316–320

4. Theunissen TW, Powell BE, Wang H et al (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell 15:471–487

5. Eiges R, Schuldiner M, Drukker M et al (2001) Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. Curr Biol 11:514–518

6. Zwaka TP, Thomson JA (2003) Homologous recombination in human embryonic stem cells. Nat Biotechnol 21:319–321

7. Davis RP, Costa M, Grandela C et al (2008) A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nat Protoc 3:1550–1558

8. Zhu Z, Verma N, Gonzalez F, Shi ZD, Huangfu D (2015) A CRISPR/Cas-mediated selection-free knockin strategy in human embryonic stem cells. Stem Cell Rep 4:1103–1111

9. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14:49–55

10. Ran FA, Hsu PD, Wright J et al (2013) Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8:2281–2308

11. Urnov FD, Rebar EJ, Holmes MC et al (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11:636–646

12. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

13. Jinek M, East A, Cheng A et al (2013) RNA-programmed genome editing in human cells. Elife 2:e00471

14. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31:230–232

15. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823

16. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. Science 339:823–826

17. Byrne SM, Ortiz L, Mali P et al (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res 43:e21

18. Hou Z, Zhang Y, Propson NE et al (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc Natl Acad Sci U S A 110:15644–15649

19. Merkert S, Wunderlich S, Bednarski C et al (2014) Efficient designer nuclease-based homologous recombination enables direct PCR screening for footprintless targeted human pluripotent stem cells. Stem Cell Rep 2:107–118

20. Merkle FT, Neuhausser WM, Santos D et al (2015) Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus. Cell Rep 11:875–883

21. Gonzalez F, Zhu Z, Shi ZD et al (2014) An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell 15:215–226

22. Zhu Z, Gonzalez F, Huangfu D (2014) The iCRISPR platform for rapid genome editing in human pluripotent stem cells. Methods Enzymol 546:215–250

23. Chen G, Gulbranson DR, Hou Z et al (2011) Chemically defined conditions for human iPSC derivation and culture. Nat Methods 8:424–429

24. Ludwig TE, Bergendahl V, Levenstein ME et al (2006) Feeder-independent culture of human embryonic stem cells. Nat Methods 3:637–646

25. Veres A, Gosis BS, Ding Q et al (2014) Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell 15:27–30

26. Borjigin J, Nathans J (1994) Insertional mutagenesis as a probe of rhodopsin's topography, stability, and activity. J Biol Chem 269:14715–14722

27. Grote E, Hao JC, Bennett MK, Kelly RB (1995) A targeting signal in VAMP regulating transport to synaptic vesicles. Cell 81:581–589

28. Nieminen M, Tuuri T, Savilahti H (2010) Genetic recombination pathways and their application for genome modification of human embryonic stem cells. Exp Cell Res 316:2578–2586

Chapter 10

Complete Transcriptome RNA-Seq

David F.B. Miller, Pearlly Yan, Fang Fang, Aaron Buechlein, Karl Kroll, David Frankhouser, Cameron Stump, Paige Stump, James B. Ford, Haixu Tang, Scott Michaels, Daniela Matei, Tim H. Huang, Jeremy Chien, Yunlong Liu, Douglas B. Rusch, and Kenneth P. Nephew

Abstract

RNA-Seq is the leading technology for analyzing gene expression on a global scale across a broad spectrum of sample types. However, due to chemical modifications by fixation or degradation due to collection methods, samples often contain an abundance of RNA that is no longer intact, and the capability of current RNA-Seq protocols to accurately quantify such samples is often limited. We have developed an RNA-Seq protocol to address these key issues as well as quantify gene expression from the whole transcriptome. Furthermore, for compatibility with improved sequencing platforms, we use restructured adapter sequences to generate libraries for Illumina HiSeq, MiSeq, and NextSeq platforms. Our protocol utilizes duplex-specific nuclease (DSN) to remove abundant ribosomal RNA sequences while retaining other types of RNA for superior transcriptome profiling from low quantity input. We employ the Illumina sequencing platform, but this method is described in sufficient detail to adapt to other platforms.

Key words RNA-Seq, Transcriptome, Gene Expression, Duplex-specific Nuclease, Sequencing

1 Introduction

RNA-seq methodology used for transcriptome analysis has been available for less than a decade [1–3], and there continues to be a need in the field for approaches capable of quantifying the complete transcriptome, regardless of size and structure. By selecting for poly(A+) RNA, standard protocols forgo completeness by utilizing random priming [4], while neglecting small RNA (smRNA) such as miRNA [5, 6]. Furthermore, RNA-seq methods for stranded sequencing approaches to capture smRNA show significant bias [7]. The protocol described in this chapter is designed to generate stranded quantitative sequence data for the entire transcriptome, which encompasses RNA sizes ranging from 20 to 20,000 nt [8]. The protocol can be used for intact, degraded, and FFPE RNA samples. Previous versions of this protocol have been

improved upon in order to allow sequencing on newer Illumina platforms (HiSeq, NextSeq, and MiSeq), incorporate low quality RNA and improve rRNA removal by duplex-specific nuclease (DSN) [9, 10]. In order to capture complete transcriptome data, we size fractionate total RNA and generate parallel complementary libraries for both small RNA (smRNA <200 nt) from native RNA and fragmented large RNA (FLgRNA >200 nt)) from total RNA. Alternatively, we describe how to generate fragmented total RNA (FTRNA) libraries without size fractionation, which is well suited for non-fragmented degraded and formalin-fixed paraffin-embedded (FFPE) samples. Adapter ligations to ssRNA preserve strandedness, which is necessary for identifying intragenic transcripts. Following cDNA synthesis, libraries are converted to dsDNA during the PCR1 step, which works with low (100 ng) quantities of starting material and suitable for qPCR analysis in place of traditional qPCR protocols that require μg input quantities [10]. Custom Illumina compatible adapters and primers are designed to allow efficient removal of adapter dimers and control duplex DNA hybrid stability during rRNA removal. High copy rRNA sequences are removed using DSN technology [9–11], which is based on hybridization kinetics and stringencies that do not target abundant adapter sequences at the ends of library inserts to be preserved for sequencing (Tables 1 and 2). The PCR2 step then completes library ends with barcodes for multiplexed sequence reads with Next Generation Sequencing. The chapter concludes with a description of library quantifications and pooling for multiplex sequencing on the Illumina platforms. The general workflow for this method is shown in Fig. 1 with a detailed library structure description in Fig. 2. This approach is described in sufficient detail to apply on other sequencing platforms.

2   Materials

2.1 RNA Isolation and Fragmentation

1. Tissue culture cells or other suitable source for RNA.
2. Microcentrifuge.
3. 1.7 mL tubes.
4. RNeasy Mini AllPrep kit (Qiagen).
5. Ethanol (100%).
6. 2-mercaptoethanol (BME).
7. Qiashredder (Qiagen).
8. RWT buffer (Qiagen).
9. Bullet Blender Storm 25 (NextAdvance).
10. 1.5 mL Rhino tubes (MidSci).
11. Magnetic 1.5 mL tube stand.

Table 1
Custom TruSeq adapters, primers and qPCR primer sequences

Assay	Adapter/Primer	Sequence
c-TruSeq Oligos[a,b]	TruSeq c-5′ adapter/ PCR1 primer	G*TTCAGAGTTCTACAGTCCGACGATC
	TruSeq c-3′ adapter	/5rApp/TGGAATTCTCGGGTGC/3ddC/
	TruSeq c-RT/PCR1-3′ primer	G*GCACCCGAGAATTCCA
	TruSeq PCR2 5′ primer (RP1)	A*ATGATACGGCGACCACCGAGATCTACACG TTCAGAGTTCTACAGTCCGA
	TruSeq PCR2 I-3′ primer (RPI#)	C*AAGCAGAAGACGGCATACGAGAT**NNNNNN** GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
qPCR Primers	5S rRNA forward primer	CGATCTCGTCTGATCTCGGAAG
	5S rRNA reverse primer	AGGCGGTCTCCCATCCAAG
	U6 snRNA forward primer	CGCTTCGGCAGCACATATAC
	U6 snRNA reverse primer	TTCACGAATTTGCGTGTCAT
	28S rRNA forward primer	TCAGACCCCAGAAAAGGTGTTG
	28S rRNA reverse primer	TGATTCGGCAGGTGAGTTGTTAC
	EEF1A1 forward primer	AAGTCTGGTGATGCTGCCATTG
	EEF1A1 reverse primer	CCCAAAGGTGGATAGTCTGAGAAG

*, phosphorothioate bond, /5rApp/, 5′-adenylation; /3ddC/, 3′-dideoxy-C, bold **NNNNNN**, the location of the reverse complement of the index sequence

12. Nuclease-free water.

13. RNeasy Minelute kit (Qiagen).

14. RNase-Free DNase kit (Norgen).

15. NanoDrop 2000 (ThermoFisher).

16. 10× fragmentation reagent and stop solution (Ambion).

17. 2100 Bioanalyzer (Agilent).

18. RNA 6000 Pico assay kit (Agilent).

2.2 End Repair and Adapter Ligation

1. smRNA or FLgRNA in 17 μL nuclease-free water.

2. 0.2 mL PCR tubes.

3. Thermocycler with a heated lid.

4. Antarctic phosphatase (New England Biolabs).

5. RNasin Plus (Promega).

Table 2
Hybridization stringencies for customized Illumina oligos for small RNA TruSeq libraries

Custom smRNA v1.5 Oligos	T_m-DSN HYB	T_m-PCR	%GC
TruSeq c-5′-adapter/PCR1 primer (same)	50.9		50.0
TruSeq c-5′-adapter/PCR1 primer overlap		62.0	47.6
TruSeq 5′ end (final)	61.1		50.9
TruSeq c-3′ adapter	48.4		58.8
TruSeq c-RT/PCR1-3′ primer	48.4	65.3*	58.8
TruSeq c-3′ adapter/TruSeq PCR1 c-3′		62.7	
primer overlap			58.8
TruSeq PCR2-5′ primer (RP1)			
TruSeq PCR2 I-3′ primer (RPI#)	64.1		53.1
TruSeq c-3′/TruSeq PCR2 I-3′ overlap		72.3	54.5
TruSeq 3′ end (final)	64.1		52.4

RT-PCR targets	T_m-DSN HYB		%GC
5S rRNA	76.4		59.5
28S rRNA	77.2		69.4

T_m, the temperature at which half of the complementary nucleic acid strands are double stranded under the salt conditions described. PCR1 library ends were customized to avoid cutting by duplex-specific nuclease (DSN) during removal of rRNA sequence inserts at 68 °C in 20 mM NaCl. T_m calculations were made using the IDT OligoAnalyzer 3.1 [11] with the appropriate oligo and salt concentrations. The DSN hybridization buffer contains 20 mM NaCl. Phusion PCR (PCR1 and PCR2) contains 50 mM KCl and 2 mM $MgCl_2$. *, SuperscriptIII RT buffer contains 50 mM KCl and 10 mM $MgCl_2$ (for the c-RT/PCR1-3′ primer in the RT reaction). The DSN hybridization buffer was adjusted to 20 mM NaCl to allow rRNA sequences to become double-stranded at 68 °C ($T_m \geq 76$ °C) while protecting the common library ends generated in PCR1 as single-strands ($T_m \leq 51$ °C)

6. T4 polynucleotide kinase.

7. 10 mM ATP (Epicentre).

8. RNase-free water.

9. RNeasy Minelute kit (Qiagen).

10. T4 RNA Ligase 2, truncated KQ (New England Biolabs).

11. TruSeq c-3′ adapter (5 μM) (*see* **Note 1** and Table 1).

12. TruSeq c-5′ adapter/PCR1 primer (10 μM) (*see* **Note 1** and Table 1).

13. T4 RNA Ligase 1 (New England Biolabs).

2.3 Reverse Transcription and PCR1 Amplification

1. 0.2 mL PCR tube.

2. Adapter ligated RNA in 10 μL RNase-free water.

3. TruSeq c-RT/PCR1-3′ primer (25 μM) (*see* **Note 1** and Table 1).

4. Superscript RT III (Life Technologies).

5. RNase H (Life Technologies).

CompleteTranscriptome RNA-seq
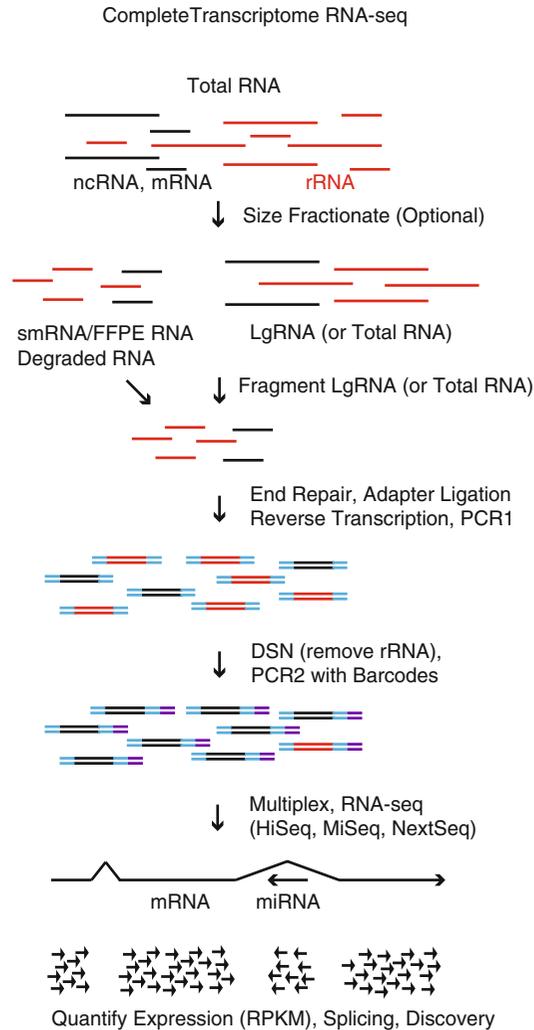
Fig. 1 General workflow for RNA-Seq libraries. Abbreviations: *DSN* Duplex-specific nuclease, *smRNA* small RNA <200 nt, *FLgRNA* fragmented large RNA >200 nt, *RPKM* Reads Per Kb per Million reads (modified from Miller et al., 2015)

6. 250 mM EDTA, pH 8.0.
7. TruSeq c-5′ adapter/PCR1 primer (25 μM) (*see* **Note** 1 and Table 1).
8. 100 mM dNTP mix.
9. Phusion DNA polymerase (New England Biolabs).
10. Thermocycler with a heated lid.
11. DNA Clean and Concentrator kit (Zymo).
12. Qubit dsDNA HS assay kit (Life Technologies).
13. Bioanalyzer HS DNA kit (Agilent).
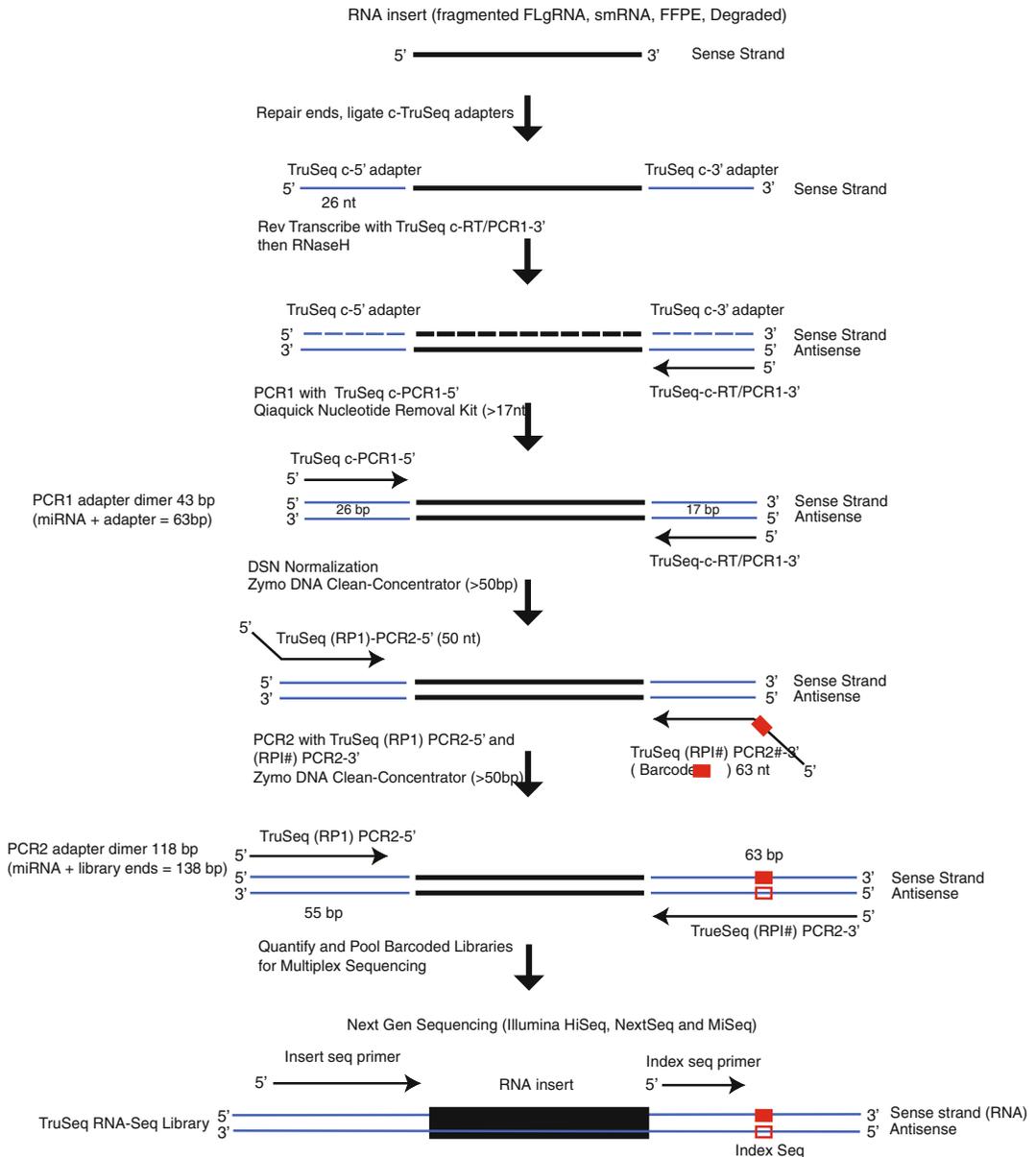
Fig. 2 Detailed description of library construction steps. This figure shows a detailed step-by-step library sequence modifications with adapter and primer locations and sizes including potential adapter dimers (modified from Miller et al., 2013)

2.4 Duplex Specific Nuclease (DSN) Reduction of rRNA Sequences

1. Nuclease-free water.
2. Thermocycler (2 chamber or additional thermocycler).
3. Duplex-specific nuclease (Wako Chemicals).
4. Custom DSN storage buffer: 28 mM Tris–HCl, 22 mM Tris base (pH 8.0 automatically, *see* **Note 2**).

5. Custom 10× DSN hybridization buffer : 200 mM NaCl, 280 mM Tris–HCl, 220 mM Tris base, 30 % vol/vol PEG8000 (40 % stock, Sigma) (pH 8.0 automatically, *see* **Note 2**).

6. 2× custom DSN reaction buffer: 28 mM Tris–HCl, 22 mM Tris base, 20 mM $MgCl_2$ (pH 8.0 automatically, *see* **Note 2**).

7. 2× custom DSN stop buffer: 500 mM NaCl, 200 mM EDTA, pH 8.0.

8. Qiagen Qiaquick Nucleotide removal kit for smRNA libraries or Zymo DNA Clean and Concentrator for FLgRNA libraries.

2.5 PCR2 Amplification with Barcode

1. 0.2 mL PCR tube.

2. DSN normalized library.

3. TruSeq c-3′ PCR2 barcoded primer (25 μM) (*see* **Note 1** and Table **1**).

4. TruSeq c-5′ PCR2 primer (25 μM) (*see* **Note 1** and Table **1**).

5. 100 mM dNTP mix (Bioline).

6. Phusion DNA polymerase (New England Biolabs).

7. Thermocycler with a heated lid.

8. AMPure XP beads (Beckman Coulter).

9. EB elution buffer (Qiagen).

10. Qubit dsDNA HS assay (Life Technologies).

11. 2100 Bioanalyzer (Agilent).

12. Bioanalyzer HS DNA kit (Agilent).

2.6 qPCR Validation of Library rRNA Sequence Reduction

1. PCR1 library.

2. DSN normalized PCR2 library.

3. 480 LightCycler (Roche).

4. LightCycler 480 SYBR Green I Master (Roche).

5. EB buffer (Qiagen).

6. Tween 20 (Promega).

7. 5S rRNA qPCR primers (*see* **Note 1** and Table **1**).

8. 28S rRNA qPCR primers (*see* **Note 1** and Table **1**).

9. EEF1A1 qPCR primers (*see* **Note 1** and Table **1**).

10. U6 snRNA qPCR primers (*see* **Note 1** and Table **1**).

11. Nuclease-free water.

12. 480 LightCycler 96-well plate (Roche).

2.7 Quantification and Pooling of RNA-seq Libraries for Multiplex RNA-seq

1. Qubit dsDNA HS assay (Life Technologies).

2. 2100 Bioanalyzer (Agilent).

3. Bioanalyzer HS DNA kit (Agilent).
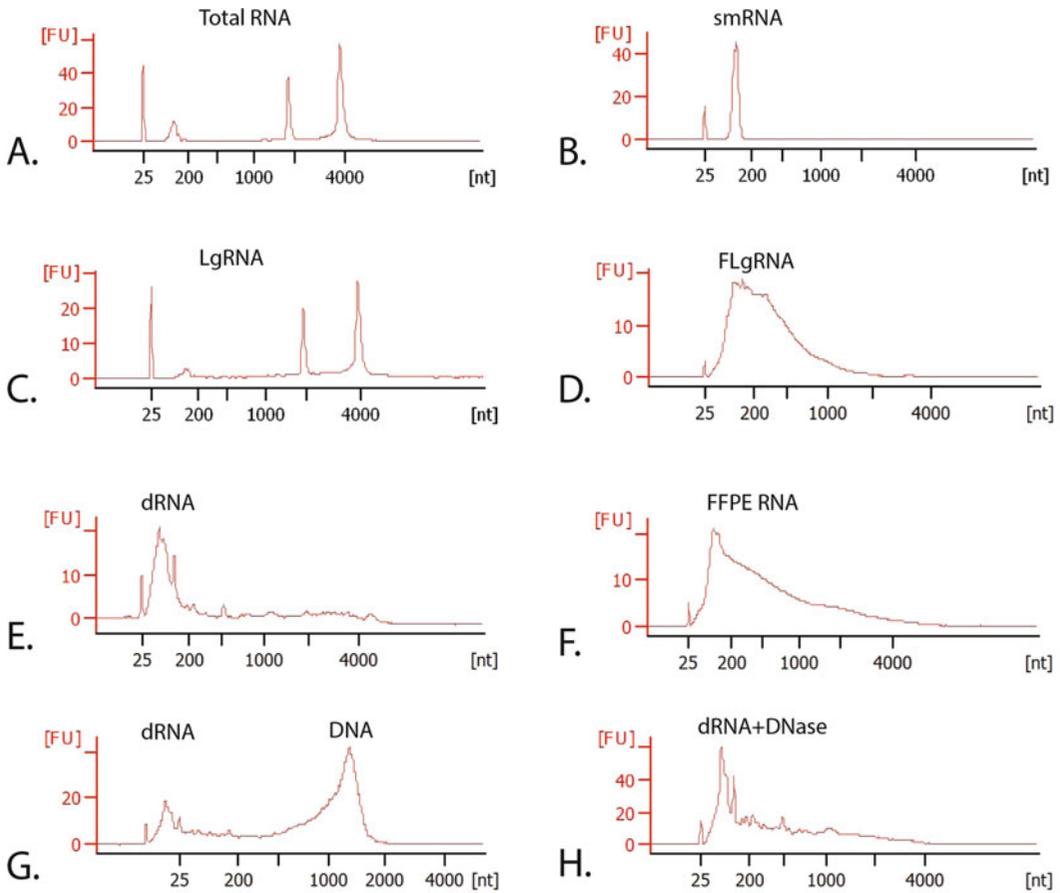
Fig. 3 Bioanalyzer RNA pico chip profiles for input RNA. (**a**) Total RNA (TRNA), (**b**) small RNA (smRNA), (**c**) large RNA (LgRNA), (**d**) fragmented large RNA (FLgRNA), (**e**) unfragmented degraded RNA (dRNA), (**f**) Formalin-fixed paraffin-embedded (FFPE) RNA, (**g**) DNA contaminated dRNA (DNA/dRNA), (**h**) DNase treated DNA/dRNA

4. EB buffer (Qiagen).

5. Tween 20.

6. Nuclease-free water.

3   Methods

3.1   RNA Purification   RNA-seq libraries can be generated from total RNA (TRNA), FFPE RNA, degraded RNA (dRNA) or TRNA size fractionated into small RNA <200 nt (smRNA) and large RNA >200 nt (LgRNA). Size fractionation of high quality RNA prior to library is necessary to sequence the complete transcriptome (*see* **Notes 3**, **4**, and **5**). Always wear gloves and perform all steps at room temperature unless specified otherwise. Purified RNA should be stored at –80 °C or on ice between manipulations. Typical size distributions for RNA types and fractions are shown in Fig. 3.

1. Add 600 µL RLT buffer containing 1 % BME (vol/vol) to the pelleted cells ($\sim 10^7$) from one 60 cm^2 culture dish and transfer to a 1.7 mL tube.

2. Rake the tube across a rack approximately 20 times or vortex the sample until the cells are completely lysed.

3. (Solid tumor only) Lyse 25 mg of solid tumor in 600 µL RLT + BME solution using a Bullet blender 24 with 1.5 mL Rhino tubes.

4. (Solid tumor only) Homogenize at speed 10 for 5 min twice. Chill sample on ice before each homgenization.

5. (Solid tumor only) Place tube on magnetic stand for 5 min to remove residual metal dust and transfer sample to a clean tube before proceeding with the standard protocol (**step 6**).

6. Aliquot 300 µL of the lysed sample and store at –80 °C indefinitely.

7. (Optional) Centrifuge 300 µL of lysed sample through a Qiashredder column for 2 min at $17,000 \times g$. Resuspend the flow-thru thoroughly.

8. Apply 300 µL of Qiashredder treated sample to the AllPrep AP DNA column and centrifuge for 20 s at $10,000 \times g$. DNA can be collected from the same sample from this column according to the manufacturer's protocol if needed.

9. Add 450 µL 100 % ethanol (60 % final vol/vol) to the AP DNA column flow-through and bind RNA to RNeasy Mini column by centrifugation for 15 s at $10,000 \times g$. The flow-through can be stored at –80 °C for protein collection according to the manufacturer's recommendations.

10. Wash the RNeasy column (centrifuge at $10,000 \times g$ for 15 s) with 500 µL RWT buffer containing ethanol according to the manufacturer (*see* **Note 3**).

11. Wash the RNeasy column (centrifuge at $10,000 \times g$ for 15 s) with 500 µL RPE buffer containing ethanol according to the AllPrep kit instructions.

12. Wash the RNeasy column (centrifuge at $17,000 \times g$ for 2 min) with 500 µL freshly prepared 80 % ethanol (vol/vol) diluted with nuclease-free water.

13. In a clean 1.7 mL tube, centrifuge the RNeasy column at $17,000 \times g$ for an additional 2 min to dry.

14. Transfer RNeasy column to a clean collection tube and allow it to dry with the lid open for at least 10 min (*see* **Note 6**).

15. Add 50–100 µL nuclease-free water to the RNeasy column, incubate at room temperature for 1 min and collect RNA by centrifugation at $17,000 \times g$ for 1 min. Resuspend the elution

thoroughly. Add a second elution for high concentrations of expected RNA yields over 30 μg.

16. Add 10 μL New England Biolabs 10× enzyme incubation buffer and 2.5 μL RNase-free DNase to each 87.5 μL volume of RNA sample. Incubate the sample at 20–25 °C for 25 min (*see* **Note 7**).

17. Add 3.5 volumes (350 μL) RLT buffer and 1.5 volumes (675 μL) 100 % ethanol to 100 μL DNase-treated RNA sample (*see* **Note 4**).

18. Purify the RNA sample with the RNeasy Mini column as above (**steps 9–15**).

19. Determine the RNA concentration with the NanoDrop (*see* **Note 8**).

3.1.2 Size Fractionation of Total RNA (Optional)

Size fractionation into smRNA (<200 nt) and LgRNA (>200 nt) is required for a complete survey of the transcriptome by sequencing in parallel (*see* **Note 4**).

1. Transfer 1–5 μg total RNA to a new 1.7 mL tube and add nuclease-free water for a final volume of 100 μL. Store additional RNA at –80 °C.

2. Add 350 μL RLT with 1 % BME and 302 μL ethanol (final 40 % vol/vol) to the RNA sample.

3. Centrifuge over an RNeasy Minelute column for 15 s at $10,000 \times g$ to bind the large RNA fraction >200 nt (LgRNA). The small RNA (smRNA) will flow through the column.

4. Add 375 μL ethanol to the smRNA fraction (final 60 % vol/vol) and bind to a clean RNeasy Minelute column by centrifugation for 15 s at $10,000 \times g$. Discard the flow-through.

5. Wash the RNeasy columns containing smRNA and LgRNA (centrifuge at $10,000 \times g$ for 15 s) with 500 μL RPE buffer containing ethanol according AllPrep kit instructions.

6. Wash the RNeasy columns (centrifuge at $17,000 \times g$ for 2 min) with 500 μL freshly prepared 80 % ethanol (vol/vol) diluted with nuclease-free water.

7. Transfer the columns to clean 1.7 mL centrifuge tubes and spin at $17,000 \times g$ for an additional 2 min to dry.

8. Transfer RNeasy columns to clean collection tubes and allow them to dry with the lid open for at least 10 min (*see* **Note 6**).

9. Add 40 μL nuclease-free water to the columns. For low concentration samples use less water.

10. Incubate the columns at room temperature for 1 min and centrifuge at $17,000 \times g$ for 1 min. Resuspend the elution thoroughly.

11. Determine the RNA concentration with the NanoDrop (*see* **Note 8**).

3.1.3 Fragmentation of Large RNA (LgRNA) or Total RNA (TRNA)

Average RNA-Seq library inserts should be fragmented to around 200 nt average size (*see* **Notes 4** and **5**).

1. Starting with 0.4–1 µg of large RNA (LgRNA) or total RNA (TRNA) in 36 µL add 4 µL 10× fragmentation reagent. This should provide adequate quantities of fragmented LgRNA (FLgRNA or FTRNA); however, lesser starting quantities can be used (*see* **Note 9**).

2. Incubate the mix at 70 °C for 4 min.

3. Stop the reaction with the addition of 4 µL 10× stop buffer to the fragmentation reaction. This should generate an average size of 150–200 nt.

3.1.4 Purify RNA with RNeasy Minelute

The RNeasy Minelute column uses the same purification method as the RNeasy Mini column but is designed for smaller quantities. FLgRNA and smRNA use different ethanol concentrations for size selection.

1. Prepare aqueous RNA for RNeasy MinElute binding by adding the following components: 44 µL RNA, 56 µL nuclease-free water, 350 µL RLT containing 1% BME, 675 µL ethanol for smRNA (final 60% ethanol vol/vol) or 302 µL ethanol for FLgRNA or FTRNA (final 40% ethanol).

2. Bind the RNA to an RNeasy Minelute column by sequential centrifugations for 15 s at $10,000 \times g$. Discard the flow-through.

3. To the column add 500 µL RPE buffer containing ethanol according to the manufacturer and wash by centrifugation for 15 s at $10,000 \times g$.

4. Add 500 µL 80% ethanol and wash the RNeasy Minelute column by centrifugation for 2 min at $17,000 \times g$.

5. Transfer the column to a clean 1.7 mL centrifuge tube and spin at $17,000 \times g$ for an additional 2 min to dry.

6. Transfer RNeasy column to a clean 1.7 mL collection tube and allow it to dry with the lid open for 10 min or more (*see* **Note 6**).

7. Add 19 µL RNase-free water to the FLgRNA columns.

8. Incubate the columns at room temperature for 1 min.

9. Centrifuge the column at $17,000 \times g$ for 1 min to elute the RNA. Resuspend the elution thoroughly.

10. Quantify the RNA fractions with the NanoDrop (*see* **Note 8**) and run 5 ng of each on the Bioanalyzer using the RNA 6000 p kit according to the manufacturer's recommendations. Typical RNA profiles are shown in Fig. 3. Check the Total RNA, smRNA, FLgRNA, and LgRNA fractions with the Bioanalyzer RNA Pico kit for incomplete size fractionation.

3.2 End Repair and Adapter Ligations

The RNeasy MinElute cleanup binding step utilizes 60% ethanol to avoid loss of smaller fragments.

3.2.1 End Repair

1. Combine the following reagents in a 0.2 mL PCR tube for phosphatase treatment of smRNA, FLgRNA, dRNA, or FFPE RNA: 17 μL RNA in RNase-free water, 2 μL 10× Antarctic phosphatase reaction buffer, 1 μL Antarctic phosphatase, 0.5 μL RNasin Plus.

2. Incubate the reaction in a thermocycler under the following conditions: 37 °C for 30 min, 65 °C for 5 min (inactivates the phosphatase), then 4 °C hold.

3. Mix the following reagents on ice for the kinase reaction: 5 μL 10× T4 polynucleotide kinase buffer, 5 μL 10 mM ATP, 2 μL T4 polynucleotide kinase, 0.5 μL RNasin Plus, and 17 μL nuclease-free water.

4. Add the T4 kinase reaction mix to the phosphatase-treated RNA in the thermocycler.

5. Incubate the reaction for 60 min at 37 °C.

6. Prepare aqueous RNA for RNeasy MinElute binding by adding the following components: 50 μL end repaired RNA, 50 μL nuclease-free water, 350 μL RLT containing 1% BME, and 675 μL ethanol (smRNA, dRNA, or FFPE RNA) or 302 μL (FLgRNA or FTRNA).

7. Bind the RNA to an RNeasy Minelute column by sequential centrifugations for 15 s at $10,000 \times g$. Discard the flow-through.

8. To the column add 500 μL RPE buffer containing ethanol according to the manufacturer and wash by centrifugation for 15 s at $10,000 \times g$.

9. Add 500 μL 80% ethanol and wash the RNeasy Minelute column by centrifugation for 2 min at $17,000 \times g$.

10. Transfer the column to a clean 1.7 mL centrifuge tube and spin at $17,000 \times g$ for an additional 2 min to dry.

11. Transfer RNeasy column to a clean 1.7 mL collection tube and allow it to dry with the lid open for 10 min or more (*see* **Note 6**).

12. Add 16 μL RNase-free water to the Minelute columns.

13. Incubate the columns at room temperature for 1 min.

14. Centrifuge the column at $17,000 \times g$ for 1 min to elute the RNA. Resuspend the elution thoroughly.

15. Quantify the RNA with the NanoDrop (*see* **Notes 6** and **8**).

3.2.2 Adapter Ligations

1. Combine these components in a 0.2 μL PCR tube to denature and ligate the 3′ adapter with end-repaired smRNA, FLgRNA, FTRNA, dRNA, or FFPE RNA (*see* **Notes 9** and **10**): 14 μL end-repaired RNA (5 pmol) and 1 μL (5 pmole) of TruSeq c-3′ adapter (5 μM).

2. Heat denature the sample for 2 min at 70 °C and hold at 4 °C.

3. Ligate the 3′ adapter in the following reaction mixture: 2 μL 10× T4 RNA ligase 2 truncated KQ buffer, 1 μL RNasin Plus, and 2 μL T4 RNA ligase 2 truncated KQ (200 u/μL).

4. Incubate the reaction overnight at 16 °C in the thermocycler.

5. Heat denature the TruSeq c-5′ adapter (10 μM) for 2 min at 70 °C and hold at 4 °C. This can be denatured just once if it is kept cold or frozen at −20 °C.

6. Ligate the 5′ adapter in the following reaction mixture: 21 μL 3′ adapter ligation, 1 μL 10× T4 RNA ligase 1 reaction buffer, 3 μL 10 mM ATP, 1 μL TruSeq c-5′ adapter (10 μM), 2 μL RNase-free water, and 2 μL T4 RNA ligase 1 (20 U/μL).

7. Incubate the reaction in the thermocycler for 2 h at 20 °C.

8. Prepare aqueous RNA for RNeasy MinElute binding by adding the following components: 30 μL RNA, 70 μL nuclease-free water, 350 μL RLT containing 1% BME, and 675 μL ethanol (smRNA, dRNA, or FFPE RNA) or 302 μL (FLgRNA or FTRNA, *see* **Note 4**).

9. Bind the RNA to an RNeasy Minelute column by sequential centrifugations for 15 s at 10,000×*g*. Discard the flow-through.

10. To the column add 500 μL RPE buffer containing ethanol according to the manufacturer and wash by centrifugation for 15 s at 10,000×*g*.

11. Add 500 μL 80% ethanol and wash the RNeasy Minelute column by centrifugation for 2 min at 17,000×*g*.

12. Transfer the column to a clean 1.7 mL centrifuge tube and spin at 17,000×*g* for an additional 2 min to dry.

13. Transfer RNeasy column to a clean 1.7 mL collection tube and allow it to dry with the lid open for 10 min or more (*see* **Note 6**).

14. Add 11 μL RNase-free water to the Minelute column.

15. Incubate the columns at room temperature for 1 min.

16. Centrifuge the column at 17,000×*g* for 1 min to elute the RNA. Resuspend the elution thoroughly.

3.3 Reverse Transcription and PCR1 Amplification

3.3.1 Reverse Transcription

1. Combine 2.5 μL TruSeq c-RT/PCR1-3′ primer (25 μM) and 10 μL adapter ligated RNA sample in a 0.2 mL PCR tube.

2. Heat denature the mixture for 2 min at 70 °C and hold at 4 °C or place on ice.

3. Add the following Superscript RTIII components to the denatured adapter-ligated RNA and TruSeq c-RT/PCR1-3′ primer mixture: 15 μL Superscript RTIII 2× reaction buffer and 2.5 μL Superscript RTIII/RNaseOut enzyme mixture. Mix gently with a pipet on ice.

4. Incubate the reverse transcription reaction for 60 min at 50 °C and then hold the reaction at 4 °C.

5. Add 1 µL RNaseH (10 U) and incubate for 20 min at 37 °C and hold at 4 °C or store at –20 °C.

3.3.2 PCR1 Amplification

1. Add the following reagents to the reverse-transcribed cDNA reaction: 20 µL Phusion reaction buffer (5×), 2 µL TruSeq c-RT/PCR1-3′ (25 µM), 2 µL TruSeq c-PCR1-5′ (25 µM), 1 µL 100 mM dNTP mix, 1 µL 100 mM EDTA pH 8 (*see* **Note 11**), 1 µL Phusion DNA polymerase (2 U), and 44 µL RNase-free water.

2. Amplify the reaction using the following PCR conditions: 1 cycle 98 °C for 30 s, 14–16 cycles at 98 °C 10 s, 55 °C for 30 s, and 72 °C for 30 s, then 1 cycle 72 °C for 10 min and 4 °C hold. Optimal amplification cycles will yield 10–20 ng/µL in 30 µL final volume (*see* **Note 11**).

3.3.3 Zymo DNA Clean and Concentrator (DCC) Cleanup for TruSeq PCR1 Reaction

The Zymo DCC kit will capture dsDNA fragments over 50 bp. PCR1 adapter dimers are 43 bp and will be removed while 20 nt miRNA inserts with adapters will be 63 bp and will be retained (Fig. 2).

1. In a 1.7 mL tube, add five volumes (500 µL) DCC DNA binding buffer and mix.

2. Centrifuge the sample in a DCC column at $10,000 \times g$ for 30 s. Remove the flow-through with a pipet to avoid column contamination.

3. Wash the column with 200 µL DNA was buffer by centrifugation at $10,000 \times g$ for 30 s.

4. Repeat the wash step (**step 3**).

5. Air-dry the column ~5 min with the lid open and elute the DNA with 30 µL EB into a clean collection tube by centrifugation at $10,000 \times g$ for 30 s. Resuspend the elution thoroughly.

6. Add 0.6 µL 5 % Tween 20 to the PCR1 sample.

7. Determine the concentration with the Qubit fluorometer and dilute and aliquot of the sample to 0.5 ng/µL in EBT (EB with 0.1 % Tween 20) (*see* **Note 8**).

8. Run 1 µL on a Bioanalyzer HS DNA chip to confirm the absence of adapter dimers. Typical PCR1 size distributions can be seen in Fig. 4a–d.

3.4 Reduction of rRNA Sequences with Duplex-specific Nuclease (DSN)

DSN specifically cuts only double stranded nucleic acids that have hybridized in the limited 5 h hybridization due to high concentrations, specifically rRNA sequences (*see* **Notes 12** and Table 2).
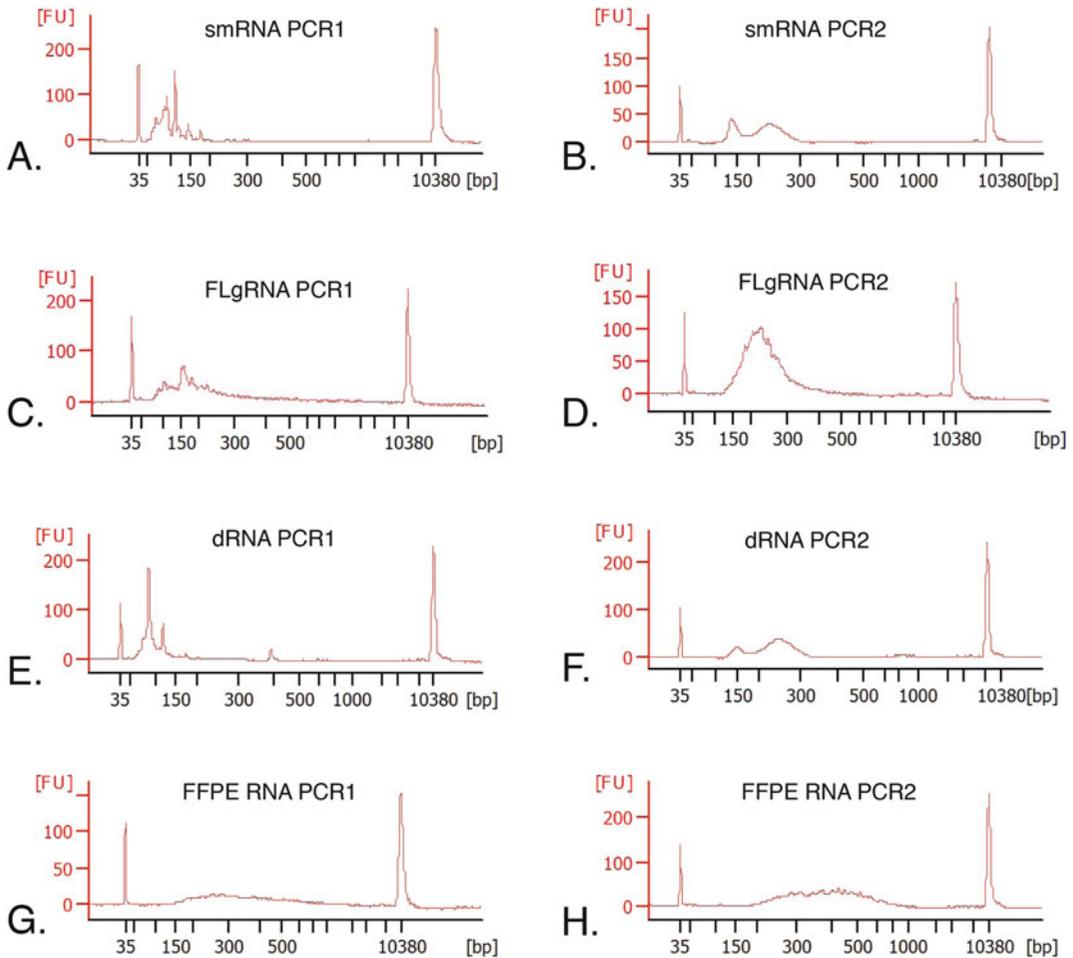
Fig. 4 Bioanalyzer HS DNA profiles for PCR1 (**A, C, E, F**) and PCR2 (**B, D, F, H**) reactions generated from (**A**, **E**) smRNA, (**B**, **F**) FLgRNA, (**C**, **G**) unfragmented degraded RNA (dRNA) and (**D**, **H**) unfragmented formalin-fixed paraffin-embedded RNA (FFPE)

3.4.1 DSN Reduction of rRNA Sequences

1. Combine the following components in a 0.2 μL PCR tube: 100 ng PCR1 library (smRNA, FLgRNA, FFPE RNA, and dRNA) or 200 ng PCR1 for FTRNA and 2 μL custom 10× DSN hybridization buffer.

2. Bring final volume to 20 μL with nuclease-free water. This will utilize a ssDNA concentration of approximately 150 nM in the hybridization reaction (*see* **Note 12**).

3. Mix the reaction with a pipette and centrifuge briefly.

4. Denature the hybridization mix at 98 °C for 2 min.

5. Incubate the hybridization for 5 h at 68 °C.

6. Preheat a 20 μL aliquot of custom 2× DSN reaction buffer in a separate thermocycler chamber at 60 °C.

7. Add 1–2 μL of the DSN enzyme to the 2× DSN reaction buffer at 60 °C and gently mix 2×.

8. Immediately add 22 μL of preheated (60 °C) 2× DSN reaction buffer/enzyme mix to the hybridization reaction at 68 °C. Gently mix the enzyme mix with the hybridization mix 10×, transferring from the bottom to the top of the solution (*see* **Note 12**).

9. Incubate the DSN enzyme reaction for 15 min at 68 °C.

10. Stop the DSN reaction by adding 41 μL of cold custom DSN 2× stop buffer and store on ice until the cleanup or store at –20 °C.

3.4.2 DSN Reaction Cleanup with QIAquick Nucleotide Removal Kit

Use the QIAquick nucleotide removal kit to retain ≥17 nt ssDNA from the DSN reaction for smRNA libraries or Zymo DNA Clean and Concentrator for FLgRNA libraries (*see* **Notes 13** and **14**).

1. Add 840 μL of buffer PNI to the DSN reaction and bind sequentially (500 μL at a time) to the QIAquick column by centrifugation at 6000 rpm ($3500 \times g$) for 60 s.

2. Remove the flow-through with a pipet and discard. Repeat **steps 1** and **2** until all the reaction is bound to the column.

3. Wash the column with 700 μL PE buffer by centrifugation at 6000 rpm ($3500 \times g$) for 60 s and discard the flow-through with a pipet.

4. Spin the column for 60 s at 13,300 rpm ($17,000 \times g$).

5. Place the column in a clean tube, air-dry for 5 min.

6. Add 50 μL EB to the column and soak for 1 min.

7. Elute the DNA by centrifugation at 13,300 rpm ($17,000 \times g$) for 60 s. Resuspend the elution thoroughly.

3.5 PCR2 Amplification with Barcode

The barcode or index sequence is added with the TruSeq PCR2-I-3′ primer (RPI#) where I indicates the unique index sequence added to each separate library [14] (Fig. 2, and Table 1) (*see* **Notes 13** and **15**).

1. Add the following reagents to the DSN normalized library: 50 μL DSN normalized library, 20 μL Phusion reaction buffer (5×), 2 μL TruSeq PCR2 I-3′ barcoded primer (25 μM), 2 μL TruSeq PCR2-5′ primer (25 μM), 1 μL 100 mM dNTP mix, 1 μL Phusion DNA polymerase (2 U), and 24 μL RNase-free water.

2. Amplify the reaction using the following PCR conditions: 1 cycle at 98 °C for 30 s, 14–16 cycles at 98 °C for 10 s, 55 °C 30 s, 72 °C 30 s, 1 cycle at 72 °C for 10 min and then 4 °C hold.

3. Check the Illumina guide for multiplexing samples [9, 14]

3.5.1 Zymo DNA Clean and Concentrator (DCC) Cleanup for TruSeq PCR2 reaction

1. In a 1.7 mL tube, add five volumes (500 μL) DCC DNA binding buffer and mix.

2. Centrifuge the sample in a DCC column at $10,000 \times g$ for 30 s. Remove the flow-through with a pipet to avoid column contamination.

3. Wash the column with 200 μL DNA wash buffer by centrifugation at $10,000 \times g$ for 30 s.

4. Repeat the wash step (**step 3**).

5. Air-dry the column ~5 min with the lid open and elute the DNA with 30 μL EB into a clean collection tube by centrifugation at $10,000 \times g$ for 30 s. Resuspend the elution thoroughly.

6. Add 0.6 μL 5 % Tween 20 to the PCR1 sample.

7. Determine the concentration with the Qubit fluorometer and dilute and aliquot of the sample to 0.5 ng/μL in EBT (EB with 0.1 % Tween 20) (*see* **Note 8**).

8. Run 1 μL on a Bioanalyzer HS DNA chip to confirm the absence of adapter dimers and determine the average size of the TruSeq library. The typical size distribution for PCR2 libraries is shown in Fig. 4e–f.

3.6 qPCR Validation of rRNA Sequence Reduction in TruSeq Libraries

The DSN normalization must be validated by qRT-PCR to ensure adequate reduction of the rRNA sequences (*see* **Note 16**).

1. Dilute PCR1 and PCR2 DNA to 0.5 ng/μL with EB buffer containing 0.1 % Tween 20 (EBT).

3.6.1 qPCR Assay for rRNA Reduction

2. Combine the following reagents in a 96-well Roche 480 LightCycler plate for qRT-PCR in triplicate: 5 μL diluted PCR1 or PCR2, 1 μL PCR primer mix (5 μM forward and reverse), 4 μL RNase-free water, and 10 μL LightCycler 480 SYBR Green I Master mix.

3. Mix each reaction (in triplicate) 5× with a pipette.

4. Use the manufacturer's recommended qRT-PCR experimental design with a 60 °C annealing temperature.

5. Analyze the rRNA reduction using the $\Delta\Delta C_T$ method [9, 10, 13] (*see* **Note 16**).

3.7 Quantify and Pool Barcoded Libraries for Multiplex Sequencing

The libraries must be accurately quantified, pooled, diluted, and quantified again before the final dilution to be sequenced [14] (*see* **Note 13**).

3.7.1 Library Quantification and Multiplexing with Barcodes

1. Determine the concentration of the libraries to be pooled using the Qubit HS DNA assay kit (*see* **Notes 8** and **15**).

2. Dilute the libraries to 0.5 ng/μL with EB buffer containing 0.1 % Tween 20 (EBT), mix and wait for 5 min to degas.

3. Run 1 μL of each diluted library on a Bioanalyzer HS DNA chip.

4. Determine the average fragment size for each library in the Bioanalyzer Region table.

5. Calculate the library concentration using the equation, $nM = \{[\text{conc (ng/μL)}] \times (10^6 \ \text{μL/L})\}/MW \ (\text{ng/nmole})$ such that $MW \ dsDNA = [(\text{size bp}) \times 607] + 158 = \text{ng/nmole}$ [9, 10] (*see* **Note 15**).

6. Dilute each library with EBT to a concentration of 17.5 nM, mix, wait for 5 min to degas, and combine equal volumes for multiplexing [9, 10, 14].

7. Determine the final concentration of the pooled libraries using the Qubit HS DNA assay kit and Bioanalyzer HS DNA chip as before.

8. Check with the Illumina instrument operator for instructions for library submission. Typically, for running on the Illumina HiSeq, add 35 fmol of the pooled libraries (2 μL of 17.5 nM) to 16.5 μL EBT with a final volume of 18.5 μL to be run in a single flow cell lane.

4   Notes

1. All adapters and PCR primers used for library generation were HPLC purified by Integrated DNA Technologies. Phosphorothioate bonds were incorporated at the 5′ base to inhibit degradation. The 3′ adapter contains a pre-adenylated 5′ end for ligation to the 3′ end of the RNA with T4 RNA ligase truncated KQ, and a dideoxy-nucleotide to avoid ligation to the 5′ end of the RNA in the second ligation step. The 3' adapter should be aliquoted and stored at -80 °C to avoid degradation. All qPCR primers we purchased from Fisher with only salt free purification. qPCR primers were designed to generate relatively small products (<100 bp) for better results with libraries constructed from degraded RNA samples. For all oligos, concentrated stocks were suspended at 1 mM concentration with TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0). Working stocks for all adapters and PCR primers were generated by diluting concentrated stocks in nuclease-free water.

2. Custom DSN (duplex-specific nuclease) buffers were generated by mixing 1 M Tris–HCl and 1 M Tris base stock solutions at the appropriate ratio (1.28:1 respectively). This produced pH 8.0 without the introduction of unnecessary salts that alter hybridization stringencies and DSN activity.

3. The Qiagen RNeasy or AllPrep kit can be supplemented with RWT buffer for the first column wash instead of the supplied RW1 wash buffer in order to retain smRNA <200 nt on the column. Following all centrifugation steps, it is necessary to thoroughly resuspend the

nucleic acids from the bottom of the tube with a pipet or high molecular weight material will be reduced.

4. RNA must be high quality (RIN > 8) for size fractionation and complete transcriptome sequencing. Total RNA can be size fractionated by changing ethanol concentrations during the RNeasy column binding steps. Ethanol concentrations at 40 % in the binding buffer will capture RNA > 200 nt while 60 % ethanol will capture all RNA size fractions. If the RNA is not reasonably high quality producing RIN values <8, then libraries should be generated from a single pool of total RNA with 60 % ethanol in the binding buffer, which can result in the loss of some smRNA transcripts when fragmented. The RIN value (RNA integrity number) can be determined by analyzing the total RNA on the Agilent Bioanalyzer RNA pico kit or a comparable assay.

5. For optimal sequencing, average fragmented library inserts should be approximately 200 nt in size. For longer reads, the fragmentation times can be reduced. Library preparations require that total RNA or LgRNA be fragmented. If RIN values from the Bioanalyzer RNA pico assay are less than RIN 5, the sample should not be fragmented (personal communication, Pearlly Yan). Go directly to the end repair step.

6. The RNeasy columns tend to retain a small amount of ethanol wash on the rim inside the column. Spinning the columns with the lid open during the dry spin step can result in cross-contamination. Remove residual ethanol from the rim with a small pipet tip and the column should be allowed to dry with the lid open on the bench to avoid a solvent peak on the NanoDrop quantification, which will alter the reported concentration and 260:280 ratio. Pure RNA should give a ratio >2.0, lower values indicate DNA, protein or solvent contamination.

7. DNase treatment is essential for accurate RNA concentration determination by the NanoDrop. DNA contamination can also be estimated by performing a Qubit HS DNA assay on the RNA samples. Typically, pure RNA samples will exhibit a concentration ratio >15 for RNA/DNA. Residual DNA will be present in the RNA fraction and promote adapter dimers due to excess adapters in the ligation steps coupled with low efficiency ligation of the 3′ adapter to DNA. These dimers and DNA can interfere with downstream applications. DNase treatment, in solution, using the New England BioLabs DNase kit performed better than other kits and resulted in higher RNA recovery than "on column" DNase treatments.

8. Use the Invitrogen Qubit fluorometer or other suitable fluorometric assay for all DNA quantifications. RNA is most accurately quantified by NanoDrop or similar spectrophotometric analysis.

9. Starting with 6 pmol of RNA is optimal for end repair reaction input given a 20 % loss during reaction cleanup steps. Based on the average sizes, this would indicate using 230 ng smRNA or 380 ng FLgRNA/FTRNA/FFPE RNA as optimal input quantities. If the starting quantities are lower, it is necessary to adjust the amount of TruSeq c-3′ adapter to be equal molar. Excess 3′ adapter will result in excess adapter dimers that can interfere with down-stream steps in the protocol. The pmole input quantity of ssRNA can be calculated with the following equations: $MW = (\#bases \times 320.5) + 159 = ng/nmole$ and $MW \, (ng/pmole) = MW/1000$.

10. The T4 RNA ligase 2 truncated KQ enzyme should only attach the pre-adenylated/dideoxy ssDNA 3′ adapter to the 3′-OH end of RNA and not DNA efficiently. However, intergenic reads above 7 % and sense strand alignments below 97 % (indicating gDNA inserts) have been observed from degraded RNA samples heavily contaminated with gDNA (personal communication, Pearlly Yan).

11. The SuperscriptIII reverse transcription reaction converts the RNA to DNA. This reaction contains 10 mM $MgCl_2$ that will compromise the Phusion DNA polymerase PCR1 amplification during PCR1. The final $MgCl_2$ concentration has been optimized with the addition of 1 μL 100 mM EDTA to the sample before the amplification in order to minimize PCR sequence bias. Adjust the number of PCR cycles to produce 10–20 ng/μL for 200 bp inserts (30 μL total). Avoid over amplification to prevent PCR biases that can produce redundant reads.

12. rRNA sequences represent about 90 % of the sequences in most RNA samples. These sequences are eliminated with duplex-specific nuclease (DSN) for efficient sequencing. Hybridizing approximately one pmole of PCR1 for 5 h allows only high copy number sequences enough time to become double-stranded once denatured. Adapter dimers, DNA contamination, salts, temperature, and the mechanics of mixing the enzyme affect DSN target specificity and activity. The salts, custom buffers and custom adapters have been designed to stabilize rRNA sequence hybrids and not the high copy library adapter sequences during the 68 °C hybridization (*see* Table 2). The 2× DSN reaction buffer is preheated to 60 °C in a second thermocycler chamber where the DSN enzyme is added and gently mixed 2× by pipetting from the bottom to the top. This 2× enzyme solution is then transferred (set the pipet at 25 μL) to the 5 h hybridization and gently mixed 10× and allowed to incubate for 15 min at 68 °C. Gentle mixing is important to retain enzyme activity. Multiple reactions can be started and stopped sequentially.

13. Sequencing depth of around 40 million reads is usually sufficient for samples generated from smRNA, FLgRNA, or FTRNA.

Adding different barcodes (index sequences) to each separate library allows the user to combine samples into one lane and then demultiplexed following sequencing. Different sequencing platforms produce different total reads per library pool so the number of libraries to be pooled will be determined by the choice of platform being utilized. Follow the Illumina guidelines for barcode combinations when multiplexing fewer than 12 samples per flow cell lane [14]. Adjust the number of PCR cycles to produce approximately 10–20 ng/μL concentrations for 200 bp inserts (30 μL total). Avoid over amplification to prevent PCR biases and redundant reads.

14. Following the DSN reduction of rRNA sequences in the library, the remaining transcriptome sequences will be ssDNA. The Zymo DNA Clean and Concentrator column will capture dsDNA 50 bp and larger but only 100 nt or larger ssDNA fragments. The Zymo DNA Clean and Concentrator column or AMPure XP beads can be used for DSN clean-up on FLgRNA libraries in section. Adapters plus 20 nt miRNA inserts are 63 nt ssDNA after DSN treatments. Consequently, smRNA containing miRNA and should be purified from the DSN reaction with the Qiagen Qiaquick Nucleotide removal kit which will capture ssDNA sizes ranging from 17 nt to 10,000 nt.

15. It is necessary to accurately quantify libraries before and after pooling. The final dilution will depend on the sequencing platform. The protocol described here is for the Illumina HiSeq platform. Quantification will require calculating the molar concentration. Average library sizes are determined with the Agilent Bioanalyzer HS DNA kit using the Region Table or other suitable assay. Library concentrations for dsDNA can be determined with the following equations: $nM = \{[\text{conc (ng/μL)}] \times (10^6 \ \mu L/L)\}/MW$ (ng/nmole) such that MW dsDNA $= [(\text{size bp}) \times 607] + 158 = \text{ng/nmole}$. Following all centrifugations, samples should be gently suspended by pipetting without introducing bubbles and allowed to degas.

16. Test rRNA sequence abundance with 5S rRNA primers for smRNA libraries and 28 s rRNA primers for FLgRNA, dRNA, or FTRNA libraries. Use the EEF1A1 primer set (or other suitable housekeeping gene) as the reference gene for the FLgRNA/dRNA/FTRNA libraries and use the U6 snRNA primers as the reference for smRNA libraries. The qPCR products should be ≤100 bp for accurate Cp values. Fold reduction can be calculated using the $\Delta\Delta C_T$ method by comparing PCR2 with PCR1 [9, 10, 13]. The 5.8S rRNA is not a suitable indicator for smRNA libraries for most RNA sources. Adequate rRNA reductions are 50-fold and 200-fold for 5S and 28S rRNA sequences respectively. It may be necessary to calculate

the C_T values with the 480 LightCycler analysis program "Absolute Quantification Analysis Using Fit Points Method" [10]. After the run, reset the upper cycle count appropriately to eliminate secondary inflection curves from interfering with the C_T calculation. PCR1 libraries can also be used to determine transcriptional differences between experimental samples for RNA quantities well below traditional qPCR assays [10].

Acknowledgements

References

1. Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

2. Nagalakshmi U, Wang Z, Waern K et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

3. Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. Cell 133:523–636

4. McGettigan PA (2012) Transcriptomics in the RNA-seq era. Curr Opin Chem Biol 17:1–8

5. Wang L, Si Y, Dedow LK et al (2011) A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-Seq. PLoS ONE 6:1–12

6. Hou Z, Jiang P, Swanson SA et al (2015) A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci Rep 5:9570–9574

7. Baran-Gale J, Erdos MR, Sison C et al. (2013) Massively differential bias between two widely used Illumina library preparation methods for small RNA sequencing. bioRxiv, http://dx.doi.org/10.1101/001479

8. Yoffe AM, Prinsen P, Gopal A et al (2008) Predicting the sizes of large RNA molecules. Proc Natl Acad Sci USA 105:16153–16158

9. Miller DFB, Yan PS, Buechlein A et al (2013) A new method for stranded whole transcriptome RNA-seq. Methods 63:126–134

10. Miller DFB, Yan PS, Fang F et al (2015) Stranded whole transcriptome RNA-Seq for all RNA types. Curr Protoc Hum Genet 84:1–11

11. Qiu X, Zhang H, Yu H et al (2015) Duplex-specific nuclease-mediated bioanalysis. Trends Biotechnol 33:180–188

12. Internet Resource. IDT (Integrated DNA Technology) OligoAnalyzer 3.1. http://www.idtdna.com/analyzer/applications/oligoanalyzer

13. Livak K, Schmitgen T (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25:402–408

14. Internet Resource. Illumina multiplexing guide. http://support.illumina.com/downloads/multiplexing_sample_prep_guide_1005361.ilmn

Chapter 11

Computational Methods and Correlation of Exon-skipping Events with Splicing, Transcription, and Epigenetic Factors

Jianbo Wang, Zhenqing Ye, Tim H. Huang, Huidong Shi, and Victor X. Jin

Abstract

Alternative splicing is widely recognized for playing roles in regulating genes and creating gene diversity. Consequently the identification and quantification of differentially spliced transcripts are pivotal for transcriptome analysis. However, how these diversified isoforms are spliced during genomic transcription and protein expression and what biological factors might influence the regulation of this are still required for further exploration. The advances in next-generation sequencing of messenger RNA (RNA-seq) have enabled us to survey gene expression and splicing more accurately. We have introduced a novel computational method, graph-based exon-skipping scanner (GESS), for de novo detection of skipping event sites from raw RNA-seq reads without prior knowledge of gene annotations, as well as for determining the dominant isoform generated from such sites. We have applied our method to publicly available RNA-seq data in GM12878 and K562 cells from the ENCODE consortium, and integrated other sequencing-based genomic data to investigate the impact of splicing activities, transcription factors (TFs) and epigenetic histone modifications on splicing outcomes. In a separate study, we also apply this algorithm in prostate cancer in The Cancer Genomics Atlas (TCGA) for de novo skipping event discovery to the understanding of abnormal splicing in each patient and to identify potential markers for prediction and progression of diseases.

Key words RNA-sequencing, Graph-based exon-skipping scanner (GESS), Alternative splicing (AS), Epigenetic

1 Introduction

Exon-skipping is the most common alternative splicing mechanism known in mammals, and is a major contributor to protein diversity in mammals. Exon-skipping results in the loss of an exon in the alternatively spliced mRNA. In this mode, the middle exon in three consecutive exons may be included in mature mRNA under some conditions or in particular tissues, but may be excluded from the mature mRNA in others. Several computational methods have been developed to detect exon-skipping events, such as ASprofile [1], DiffSplice [2], and DSGseq [3]. Notably, all of the abovementioned methods have been proven to be useful in detecting novel motifs and deciphering the logics of alternative splicing [4]. To this

end, our group has developed a novel computational method, graph-based exon-skipping scanner (GESS) [5] (detection scheme summarized in Fig. 1). Remarkably, a notable advantage of our GESS method is reflected in the capability of capturing de novo exon-skipping events from raw RNA-seq data without the prior knowledge of gene annotation information [6].

Since the mechanism of transcriptional regulation in a cell is complex and dynamic, resulting in diverse outcomes under different physiological conditions, many current approaches for the identification of skipping event depend on annotated exon information. Not only such approaches may be unable to capture the full landscape of gene expression in situ, but also sometimes may lead to errors in the interpretation of results [7]. To the contrary,
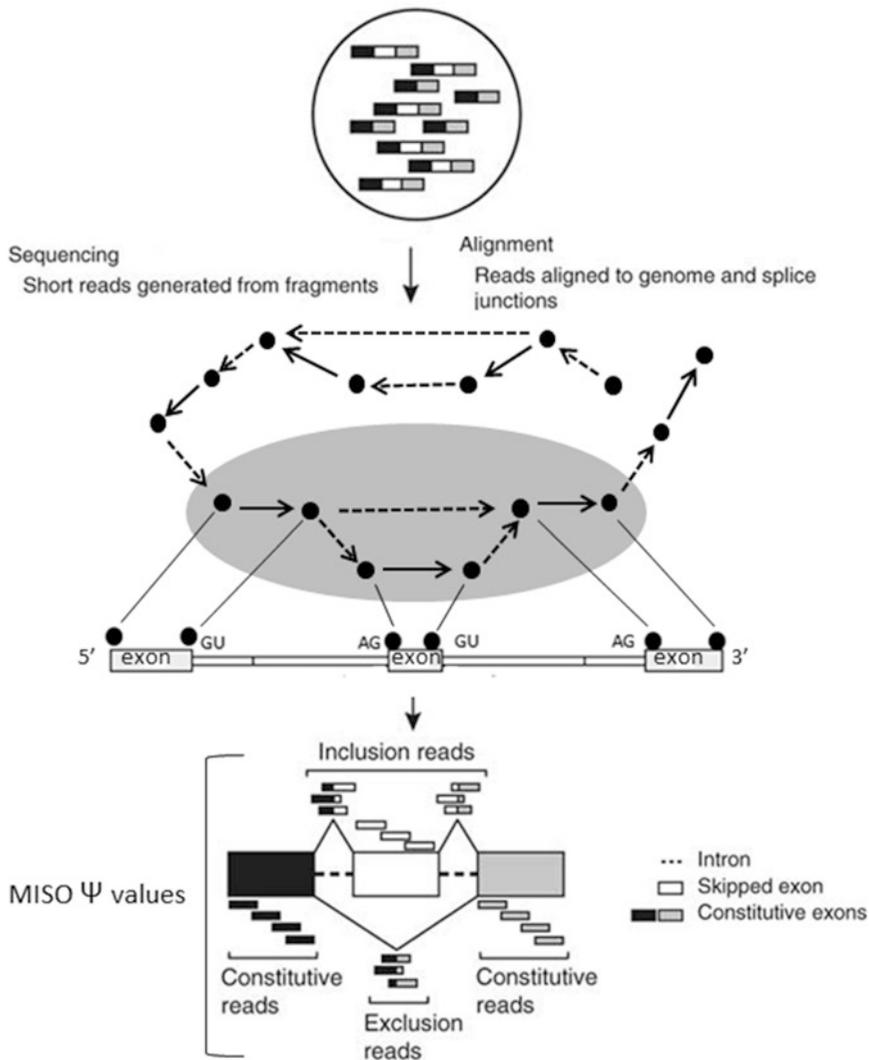
Fig. 1 The scheme of the exon-skipping event detection pipeline (GESS)

our GESS method rather builds a splice-site-link graph from first-hand, raw RNA-seq reads and then implements a walking strategy on this graph by iteratively navigating sub-graphs to reveal those with a pattern corresponding to an exon-skipping event. Thus, it can provide a more accurate and comprehensive picture of skipping events associated with a particular physiological condition within a cell. Furthermore, we integrated the MISO model into our method to determine which isoform, skipping- or inclusion-isoform, is the dominant transcript produced from a skipping-event site, where the maintenance of the subtle balance between the two mRNA molecules is indeed vital to cellular function and dynamics.

2   Methods

The flowchart in Fig. 1 exhibits the general protocol used for the discovery of de novo splicing events.

2.1   Tophat Splicing Aware Alignment

1. Input raw RNA-seq data set (either in single-end or pair-end sequences) in FASTQ or FASTA format.

2. Bad reads with low quality and ambiguous bases were filtered out.

3. Process the input data set in TopHat [8] and align the remaining reads to the reference genome (either human hg18/19 or mouse mm8/9/10).

4. Remaining set of unique aligned reads are composed of two subsets:

    (a) A set of aligned splicing-reads in which those reads are split between two genomic locations (presumably the putative exon's junction).

    (b) A set of aligned constitutive-reads in which those reads are restrictively mapped to the same genomic location without splitting two locations (presumably within one exon).

2.2   Introduction to GESS

1. Assign the two chromosome positions of a junction revealed by a spliced-read into two nodes, each corresponding to the potential splice site.

2. Link the two nodes with an edge in a dotted line if a certain number (default parameter is 5) of spliced-reads are above the defined threshold.

3. Determine the direction of the line by examining the "GT-AG" consensus rule for most vertebrate introns since the dotted-line edge corresponds to an intron gap.

4. Calculate the coverage density among these splice sites using the set of constitutive-reads.

5. Link the two splice sites (nodes) with an edge in a solid line if higher density of reads between two splice sites. This type of edge should correspond to those exonic regions.

6. Sort these splice sites along the chromosome coordinates, and calculate the depth of coverage for each segment between the two adjacent splice sites (*see* **Note 1**).

   As shown in Fig. 1 (the grey oval shape), a walking strategy on this graph by iteratively navigating the sub-graphs with pattern introduce an exon-skipping event.

7. Check the pattern which should conform to tri-exons with three solid edges, and the downstream exon would be connected to the upstream exon indicated by the dotted edges.

8. Ignore patterns that are not matched and move to next combination.

9. Define these confirmed sub-graphs as exon-skipping events with two possible combinations:

   the inclusion combination (termed as inclusion isoform)—inclusion of the middle exon

   the skipping combination (termed as skipping isoform)—exclusion of the middle exon

10. Integrate a MISO [9] model to calculate the ratio of two isoforms and determine which isoform is a dominant event in this cellular condition using the following formula.

    Ratio of two isoforms

    $$\psi f_1 = c_1 \psi_1 \ / \left(c_1 \psi_1 + c_2 \psi_2 \right)$$
    $$\psi f_2 = 1 - \psi f_1$$
    $$\psi_{\mathrm{big}} : \psi f_1 \geq 0.7 \left(\text{inclusion isoform predominates} \right)$$
    $$\psi_{\mathrm{sml}} : \psi f_1 \leq 0.3 \left(\text{skipping isoform predominates} \right)$$

    For more information on integration of MISO, please view the MISO website http://miso.readthedocs.org/en/fastmiso/.

3   Application to K562 and GM Cells (Lymphoid Origin)

To demonstrate its performance and applicability, we applied the GESS method to publicly available RNA-seq data from K562 and GM12878 cells [10]. These two cell lines are ENCODE Tier 1 cell lines with many publicly available "omics" datasets for further analysis available for each [11]. Using GESS, we identified 2750 exon-skipping events in K562 cells and 3583 events in GM12878 cells. Of these events, 1299 were common to both cell lines (Fig. 2a). Comparing our results to the annotated exon-skipping database for the human genome, which contains 39,232 events and was
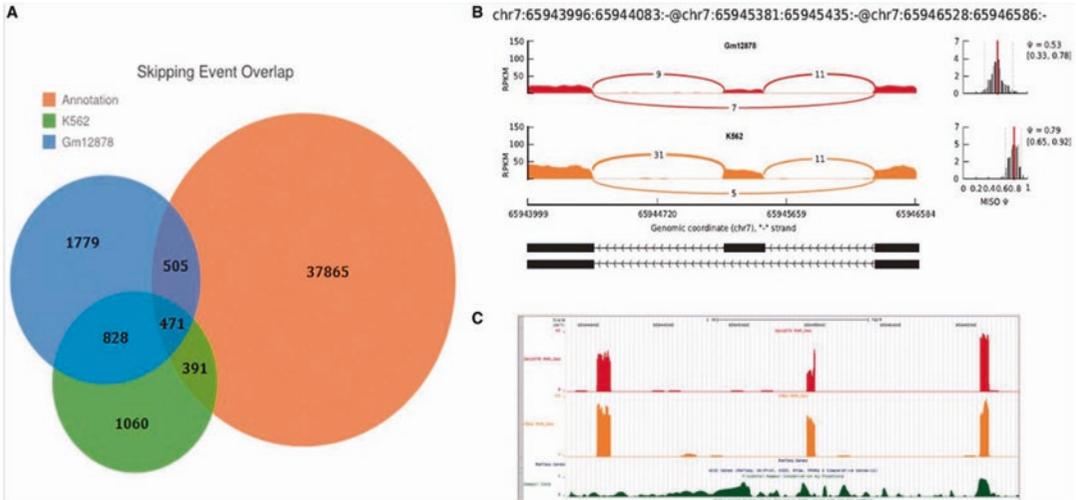
Fig. 2 (**a**) A Venn diagram showing an overlapping comparison of exon-skipping events identified by GESS with the annotated events from the MISO website. (**b**) An exon-skipping event detected by GESS, in which both isoforms are present in K562 and GM12878 cells. (**c**) No RefGene information for this skipping event was found on the UCSC track (*top panel*); the coverage along the chromosome is also provided (*bottom panel*)

downloaded from the MISO website, we found only ~30 % of our events overlapped previously annotated skipping events, with many unique skipping events being newly detected by our method. We also observed that a large amount of annotated events were not reported by GESS due to absent/low expression signals or splicing links in the RNA-seq data utilized. By comparing GESS-predicted skipping events with the annotated RefSeq database (UCSC HG19 RefSeq), in which each exon-skipping event can be mapped to a specific annotated gene, we found 40 skipping events that were not assignable to any known genes in K562 cells, while 34 events lacked annotations in GM12878 cells. As an example shown in Fig. 2b, we observed three adjacent exons on chromosome 7 covered by numerous reads in which the alignment pattern of splicing-reads revealed two isoforms with differential expression ratios in the two cell lines. However, no gene annotation information exists for this genomic region (see the RefSeq gene track in Fig. 2c) and no skipping event annotation can be found in the MISO dataset.

4   Application to PCa Patients

Prostate cancer (PCa) is the most common cancer and the second cause of cancer death among men in European countries [12]. In general, PCa is a highly heterogeneous disease, ranging from slow-growing tumors to rapidly progressing highly aggressive carcinomas associated with significant morbidity and mortality. Therefore, early detection of PCa by measuring prostate specific antigen (PSA)

values at regular intervals in peripheral blood is important to identify men with aggressive cancers at early stage [13, 14] (Table 1).

Next generation RNA sequencing data were generated by The Cancer Genomics Atlas (TCGA) consortium for 558 samples, 48 benign samples and 510 primary tumors. 96 of these samples represented advanced disease with Gleason grade ≥8 [15] (Table 2) and 33 cases had undergone progression as characterized by postoperative biochemical recurrence. Data were downloaded from the UCSC Cancer Genome Browser (Fig. 3) (https://browser.cghub. ucsc.edu/). Associated clinical data were downloaded from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga/).

Our Initial sequencing studies together with GESS illustrated previously suggest that the upregulation of selected splicing regulators in PCa, such as SAM68, SRSF1, or DDX5, directly contributes to the phenotype by altering the splicing profile of key genes [16]. The potential value of targeting specific components of the splicing machinery in cancer cells is also suggested by the antioncogenic properties of natural compounds, such as spliceostatin A (SSA), in a variety of cancer cell models. SSA targets the splicing factor 3B subunit 1 (SF3B1) of the spliceosome, thus affecting a large number of splicing events concomitantly [17]. The PCa genome appears to be characterized by rare SNP and frequent copy-number aberrations and genomic rearrangements. These rearrangements seem to arise in a punctuated manner, driving clonal expansion and evolution [18].

Table 1
Prostate cancer stage

Stage I	The tumor is small and only in the prostate
Stage II	The tumor is larger and may be in both lobes of the prostate but is still confined to the prostate
Stage III	The tumor has spread beyond the prostate to close by lymph glands or seminal vesicles
Stage IV	The tumor has spread to other organs such as the bone and is referred to as metastatic cancer. If prostate cancer spreads, or metastasizes, to the bone, one gets prostate cancer cells in the bone

Table 2
Gleason scores in categorical order

Gleason 6	The tumor tissue is well differentiated, less aggressive, and likely to grow more slowly
Gleason 7	The tumor tissue is moderately differentiated, moderately aggressive, and likely to grow but may not spread quickly
Gleason 8–10	The tumor tissue is poorly differentiated or undifferentiated highly aggressive and likely to grow faster and spread

Fig. 3 Screenshot of PCa raw RNA-seq data download page from the UCSC Cancer Genome Browser Data Portal (https://browser.cghub.ucsc.edu/)

5   Correlation of Epigenetic Marks with Exon-skipping Events

It has been widely accepted that chromatin state plays essential roles in regulating gene expression. While DNA methylation, nucleosome occupancy and modifications of histone are all involved in determining the chromatin state, some transcription factors (TFs) can bind to specific regulatory regions to interact with chromatin and regulate gene expression [19]. All these factors can be considered as epigenetic features that regulate gene expression from a broad perspective [20].

In order to understand the relationship between chromatin modifications and exon skipping events, we analyzed "omics" data for two epigenetic marks associated with transcription elongation, H3K36me3 and H3K79me2 [21]. In GM12878 cells, we found that H3K36me3 is not only involved in coupling transcription and splicing events, but also in regulating splicing processes in a cell type- and perhaps gene site-specific manner. For H3K79me2, we observed that it is enriched over splice sites in the ψ_{sml} group versus the ψ_{big} group in both cell types. Interestingly, with the exception of H3K79me2, the distribution of these transcription and epigenetic factors exhibited decreasing enrichment when progressing from an exon toward an intron. However, increasing enrichment was noted when progressing from an intron to an exon. This suggests these factors either may participate or show sensitivity to exon–intron boundary establishment. Taken together, our analysis suggests that different epigenetic factors may introduce a variable obstacle in the process of exon–intron boundary establishment leading to skipping events.

6   Note

1. For each specific segment carrying a robustly higher signal ratio (i.e., 3.0) relative to the flanking background segments, a solid edge is introduced as an exon gap. Thus a complex graph would be obtained with intronic or exonic links among the splice sites.

References

1. Florea L, Song L, Salberg SL et al (2013) Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues [v2; ref status: indexed, http://f1000r.es/2dl]. F1000Research 2:188

2. Hu Y, Huang Y, Du Y et al (2013) DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res 41:e39. doi:10.1093/nar/gks1026

3. Wang W, Qin Z, Feng Z et al (2013) Identifying differentially spliced genes from two groups of RNA-seq samples. Gene 518:164–170

4. Laurent L, Wong E, Li G et al (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20:320–331

5. Ye Z, Chen Z, Lan X et al (2014) Computational analysis reveals a correlation of exon-skipping events with splicing, transcription and epigenetic factors. Nucleic Acids Res 42:2856–2869

6. Pushkarev D, Neff NF, Quake SR (2009) Single-molecule sequencing of an individual human genome. Nat Biotechnol 27:847–850

7. Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res 35(Web Server issue):W297–W299

8. Trapnell C, Pachter L, Salberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

9. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 7:1009–1015

10. Trapnell C, Williams BA, Pertea G et al (2010) Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat Biotechnol 28:511–515

11. The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74

12. Catalona WJ, Partin AW, Finlay JA et al (1999) Use of percentage of free prostate-specific antigen to identify men at high risk of prostate cancer when psa levels are 2.51 to 4 ng/mL and digital rectal examination is not suspicious for prostate cancer: an alternative model. Urology 54:220–224

13. Jemal A, Bray F, Center MM et al (2011) Global cancer statistics. CA Cancer J Clin 61:69–90

14. Shen MM, Abate-Shen C (2010) Molecular genetics of prostate cancer: new prospects for old challenges. Genes Dev 24:1967–2000

15. Clarke RA, Schirra HJ, Catto JW et al (2010) Markers for detection of prostate cancer. Cancers (Basel) 2:1125–1154

16. Berger MF, Lawrence MS, Demichelis F et al (2011) The genomic complexity of primary human prostate cancer. Nature 470:214–220

17. Kornblihtt AR, de la Mata M, Fededa JP et al (2004) Multiple links between transcription and splicing. RNA 10:1489–1498

18. Baca SC, Prandi D, Lawrence MS et al (2013) Punctuated evolution of prostate cancer genomes. Cell 153:666–677

19. Goodrich JA, Tjian R (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet 11:549–558

20. Barski A, Cuddapah S, Cui K et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837

21. Wang ET, Sandberg R, Luo S et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476

Chapter 12

Tissue Engineering Platforms to Replicate the Tumor Microenvironment of Multiple Myeloma

Wenting Zhang, Woo Y. Lee, and Jenny Zilberberg

Abstract

We described here the manufacturing and implementation of two prototype perfusion culture devices designed primarily for the cultivation of difficult-to-preserve primary patient-derived multiple myeloma cells (MMC). The first device consists of an osteoblast (OSB)-derived 3D tissue scaffold constructed in a perfused microfluidic environment. The second platform is a 96-well plate-modified perfusion culture device that can be utilized to reconstruct several tissue and tumor microenvironments utilizing both primary human and murine cells. This culture device was designed and fabricated specifically to: (1) enable the preservation of primary MMC for downstream use in biological studies and chemosensitivity analyses and, (2) provide a high-throughput format that is compatible with plate readers specifically seeing that this system is built on an industry standard 96-well tissue culture plate.

Key words Microenvironment, Three-dimensional, Microfluidics, Perfusion, Multiple myeloma

1 Introduction

1.1 Tissue Engineering Platforms in the Era of Personalized Medicine

Most of the current knowledge in the field of cancer biology and drug discovery has had its roots in the two-dimensional (2D) culture of cell lines, in which the three-dimensional (3D) features and complex interplays between primary patient-derived tumor cells and their microenvironment have largely been overlooked. As a result, the vast majority of new anticancer drugs eventually fail in clinical trials, despite strong indications of activity in in vitro preclinical studies [1, 2]. In particular, the use of primary cells is important since immortalizing human cells perturbs the cells' gene expression profiles and cellular physiology [3–5]. Even if primary cells can be grown and maintained, resulting genomic expression and cellular physiology can be rather different in conventional versus microenvironment-mimicking culture systems. In this regard, as demonstrated over two decades ago by the seminal work of Bissel and colleagues [6–8], it is now well recognized that cancers are complex and heterogeneous and because of this the flat and hard plastic or glass

substrates commonly used for cell culture are not representative of their in vivo cellular microenvironment. Furthermore, the simplified culture conditions of 2D surfaces fall short on reproducing the mechanical and biochemical signals, as well as the cell–cell and cell–extracellular matrix (ECM) adhesive interactions which can be of critical importance not only for the ex vivo preservation of primary healthy and tumor cells, as we and other have demonstrated [9–13], but for the maintenance of stromal cues known to confer tumors with drug resistance [14–17]. Therefore, there has been a rapidly growing recognition for the need to develop new technologies that can aid in the recapitulation of physiologically relevant human tissue and tumor models to: (1) conduct preclinical drug evaluation and reduce reliance on in vitro and animal models that correlate poorly with clinical outcomes, and (2) develop patient-specific diagnostic screening platforms to evaluate personalized treatment options for the optimum care of patients [18, 19].

Aside from the need for having a suitable microenvironment-mimicking framework composed of functional and multicellular stromal elements of the tumor, it is anticipated that in vitro culture technologies aimed at recapitulating the tumor niche will also require significant advances on three major research fronts: (1) ability to work with primary human cells which are often difficult to maintain ex vivo, (2) developing and manufacturing of devices that can be easily used by laboratory personnel to evaluate cellular responses to drugs in a high-throughput manner, and (3) incorporate novel techniques for high-content screening.

On the development of static 3D tumor models, the most currently used technologies to date are: (1) embedding single cells or aggregates into a 3D biomaterial scaffold made up of collagen, alginate, or Matrigel™, (2) scaffold based culture systems, (3) monocellular and multicellular spheroids, and (4) cell printing. Each one of these techniques comes with its advantage and disadvantages [20–22], but although these systems are beneficial to maintain the 3D cellular architecture of tissues, they have limitations in replicating the mechanical forces and mass transfer conditions present within the in vivo tumor microenvironment. To this end, bioreactors culture systems have been designed to provide efficient mass transfer, computer-controlled systems provide online monitoring and automated control of environmental culture variables [23–26]. More recently, microfluidic-based perfusion culture has also emerged at the technological forefront to maintain physiological levels of environmental parameters known to influence cell behavior and consequently drug responses [10, 13, 27–30].

However, it is anticipated that for their practical and wide use, perfusion culture devices will need to: (1) allow the convenient and spatiotemporal placement of cells and biomaterials in wells, (2) be able to support long-term (for up so several weeks) multicellular culture of primary patient-derived cells, which may be required for

functional 3D tissue replication as well as monitoring long-term cell response to drugs, and (3) be compatible with conventional tissue and cell characterization techniques including phenotyping, genotyping, and viability analysis. Ultimately, devices ought to be commercially available at reasonable costs relative to those utilized in traditional in vitro cultures.

1.2 Multiple Myeloma

Multiple myeloma (MM), an incurable hematological malignancy characterized by the incontrollable proliferation of plasma cells in the bone marrow, is the second most common blood cancer in the USA with a typical survival of 5–7 years [31, 32]. It is now well known that MMC interaction with stromal elements of the tumor microenvironment is involved in tumor progression and the development of drug resistance; specifically, MMC preferential interaction with osteoblasts (OSB) at the bone marrow endosteal layer has been implicated in conferring drug resistance and facilitating the survival of malignant plasma cells [33–36]. Because of this interplay, one of the major challenges associated with studying and evaluating new therapeutic approaches against MM and other cancers that likewise reside or metastasis to the bone/bone marrow space has been the lack of clinically relevant, high-throughput, and inexpensive ex vivo models. The complex in vivo bone marrow microenvironment consists of non-hematopoietic and hematopoietic cells, together with an extracellular and liquid compartment organized in a complex architecture of sub-microenvironments (referred to as "niches") within a protective coat of mineralized bone. The balance between these different compartments is profoundly affected in multiple myeloma.

The replication of the intricate multiple myeloma niche has been examined using both in vitro systems and in vivo models. Yaccoby and Epstein developed the widely used xenograft model capable of supporting the growth of human myeloma [37, 38], using a severe combined immunodeficient (SCID) mouse implanted with human fetal bone fragments to create a humanized bone microenvironment (SCIDhu). Lawson and colleagues [39] just described a new xenograft model where NOD/SCID-GAMMA (NSG) mice were MMC injected via the tail vein homed to the bones of challenged mice resembling human disease. This simplified approach proved beneficial for the assessment of various drug treatments. However, the use of animals is overall time consuming, costly and has limited throughput capacity for drug evaluation. Also in recent years, a number of groups have developed different kinds of 3D culture systems with several key niche elements that allowed in vitro expansion of MMC [40, 41]. However, the conventional macroscale static culture nature of these approaches has not been suitable for assessing the potential contribution of perfusion-related mass transfer and shear stress effects on the dynamic progression and expansion of MMC. The interactions of MMC within their

niche, in a perfused environment that can be controlled to emulate blood flow variations due to vascularization changes and recapitulates the spatial and temporal characteristics of the complex tumor microenvironment, has yet to be incorporated and examined in in vitro models of multiple myeloma.

We described here the manufacturing and implementation of two prototype perfusion culture devices designed primarily for the cultivation of difficult-to-preserve primary patient-derived MMC, although their utility go beyond this particular application and tumor type. The first device consists of an OSB-derived 3D tissue scaffold constructed in a perfused microfluidic environment [10]. Because it is designed on a glass slide, this device is particularly suitable to conduct time-lapse microscopy. To begin to address some of the more complex challenges detailed above with regards to the broader implementation of perfusion culture systems, we also developed a 96-well plate-modified perfusion culture device for reconstructing several tissue and tumor microenvironments utilizing both primary human and murine cells [13, 42]. This culture device was designed and fabricated specifically to: (1) enable the preservation of primary MMC for downstream use in biological studies and chemosensitivity analyses and, (2) provide a high-throughput format that is compatible with plate readers specifically seeing that this system is built on an industry standard 96-well tissue culture plate.

In both systems patient-derived bone marrow mononuclear cells are coculture with the human OSB cell line hFOB 1.19, which over time develops into a tissue construct constituted by cells and their deposited ECM and provides a critical biological substrate supporting MMC.

2   Materials

2.1   Patient-Specific 3D Microfluidic Tissue Culture Device

1. Polydimethylsiloxane (PDMS, Sylgard 184 silicone elastomer kit, Dow Corning Corp.).

2. PDMS curing agent (Sylgard 184, Dow Corning Corp.).

3. Photoresist material (SU8 2025, Microchem, Newton, MA).

4. Razor blade.

5. Oxygen plasma (PDC-001, Harrick Plasma).

6. 1.2 mm-thick large glass slide (4″ × 3″, Ted Pella).

7. Hand press (Schmidt Technology).

8. Stainless steel pins (0.63 mm outer diameter, New England Small Tube).

9. Stainless steel dispensing needle (Straight, 23 Gauge, 1–1/2″ Long, McMaster-Carr).

10. Stainless steel dispensing needle (Straight, 23 Gauge, 1/2″ Long, McMaster-Carr).

11. Polyethylene tubing (0.58 mm inner diameter, Scientific Commodities).

12. Poly ε-caprolactone (PCL).

13. 1,1,1,3,3,3-hexafluoro isopropanol (HFIP).

14. Scotch tape.

15. Power voltage supplier (Gamma High Voltage Research, Inc.).

16. VOA vials open top with 0.125 in PTFE/silicone septa (40 mL, Fisher Scientific).

17. 70 % (v:v) isopropanol (IPA).

18. Disposable 50 mL centrifuge tube vacuum filter (0.22 μm polyethersulfone membrane filter, Millipore).

19. Human fibronectin (BD Biosciences).

20. Syringe pumps (New Era Pump Systems, Inc.).

21. 10 mL syringes (BD Biosciences).

22. 40 μm cell strainer (Falcon).

23. Microcentrifuge Tube (Fisher Scientific).

2.1.1 Cell Lines and Primary Cells

1. Human osteoblasts cell line hFOB 1.19 (CRL-11372, American Type Culture Collection).

2. Patient-derived bone marrow mononuclear cells (BMMC).

2.1.2 Medium Formulation and Other Solutions

Media components can be purchased from various suppliers such as GIBCO and Fisher Scientific.

1. hFOB 1.19 culture medium: 1:1 mixture of Ham's F12 medium and Dulbecco's modified. hFOB 1.19 culture medium is supplemented with 10 % fetal bovine serum (FBS) and 0.3 mg/mL geneticin (G418) (Life Technologies).

2. Bone marrow culture medium: RPMI with l-glutamine, plasma mixture (described below), 6.2×10^{-4} M of $CaCl_2$, 1×10^{-6} M sodium succinate, 1×10^{-6} M hydrocortisone, 1 % penicillin–streptomycin (p/s), 1 unit/mL heparin. The complete bone marrow medium is decontaminated using 50 mL centrifuge tube vacuum filter units (EMD Millipore).

3. Complete RPMI-1640 medium (cRPMI): RPMI-1640 medium supplemented with 10 % FBS, 1 % p/s, 1 % l-glutamine.

4. Phosphatase-buffered saline (PBS) (Sigma-Aldrich).

5. Fluorescence-activated cell sorting (FACS) Buffer: dissolve 10 g of bovine serum albumin (BSA, Sigma-Aldrich) and 0.2 g sodium azide (Sigma-Aldrich) in 1 L of 1× PBS. Mix with stir bar for 15 min and sterile filter.

6. Plasma mixture (10 % v/v) is composed of 5 % pooled plasma from 5 to 10 patients, and 5 % normal human serum type AB (AB serum) (Atlanta Biologicals).

2.1.3 Tissue Dissociation

1. Trypsin–EDTA solution (0.25 %) (GIBCO).

2. Disposable polypropylene, RNase-free pellet pestle (Thomas Scientific).

2.1.4 Antibodies and Stains

Antibodies can be purchased from BD Pharmingen, Beckman Coulter and Miltenyi Biotec.

A flow cytometer with 5-color capability (minimum) is needed to perform flow cytometric analyses.

1. Carboxyfluorescein diacetate, succinimidyl ester (CFSE) (Life Technologies).

2. Mouse anti-human CD138-APC (allophycocyanin; APC, clone B-B4).

3. Mouse anti-human CD38-PECy5 (clone LS198-4-3).

4. Mouse anti-human CD38-PECy7 (clone IB6).

5. Mouse anti-human CD56-PE-Vio770 (clone AF12-7H3).

6. Mouse anti-human CD13-PE (phycoerythrin; PE, clone WM15). CD13 is an OSB extracellular marker not expressed by BMMC.

7. 7-aminoactinomycin D (7-AAD, Life Technologies).

2.2 96-Well Plate Perfusion Culture Device

Aside from the described materials in Subheading 2.1, the preparation of the 96-well plate culture platform also requires of the following materials.

1. Polystyrene (PS) bottomless 96-well plate (Greiner Bio-One).

2. Polycarbonate (PC) membrane sheets (Sterlitech).

3. Hole puncher (6 mm, Miltex).

4. Hole puncher (3/8-in., EK Tools Circle Punch).

5. Stainless steel round punch (0.6 mm inner diameter, Technical Innovations).

3 Methods

Here we describe an approach for culturing difficult-to-preserve primary patient-derived multiple myeloma cells (MMC) using an osteoblast (OSB)-derived 3D tissue scaffold constructed in a perfused microfluidic environment and a culture medium supplemented with patient plasma.

3.1 Microfluidic Device Fabrication

As shown in Fig. 1a, a glass slide is integrated with an 8-chamber microfluidic housing made out of PDMS. As shown in Fig. 1b, each chamber had an elongated hexagonal prism shape (12 mm long, 6 mm wide, and 0.2 mm deep) with a chamber volume of ~10 μL.
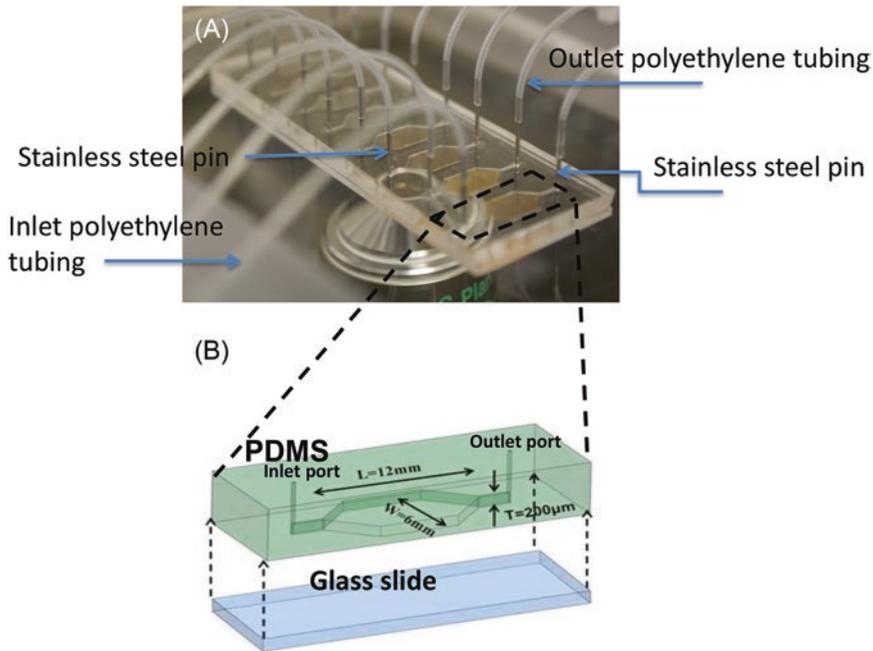
Fig. 1 Microfluidic device with eight culture chambers. (**a**) Microfluidic device can be operated on a micro-scope stage housed in an environmental chamber for in situ microscopy. (**b**) Chamber dimensions [28]. Reproduced from Lee J.H. et al., 2012, DOI: 10.1016/j.biomaterials.2011.10.036, with permission from Elsevier Publishing Company

1. Design a system of eight elongated hexagonal prism shape (12 mm long, 6 mm wide, and 0.2 mm deep) using AutoCAD (computer-aid-design program from Autodesk).

2. Use the CAD file to produce a high-resolution transparency using a commercial printer.

3. Used the transparency sheet as a photomask to produce a master. The master contains a positive pattern of eight elongated hexagonal prism shapes on a silicon wafer by photolithography with a photoresist material.

4. Mix the PDMS precursor and the curing agent at a 10:1 ratio (~30 g total).

5. Pour the mixture on the patterned silicon wafer; briefly spin at $5 \times g$ to allow the liquid to cover the round shaped silicon plate.

6. Degas the mixture in a vacuum desiccator.

7. Cure at 70 °C for 1 h to create the 300 mm thick PDMS top.

8. Peel off the PDMS housing from the mold.

9. Punch holes for each chamber through both inlet and outlet using a hand press and a stainless steel round punch (0.6 mm inner diameter) for fluidic connections.

10. Use scotch tape to cover the glass slide and only leave a 2 mm wide area (5 mm to the edge) along the longitudinal direction.

11. Make 8 % w/v PCL solution in 1,1,1,3,3,3-hexafluoro isopropanol (HFIP) as electrospinning solution.

12. Transfer the electrospinning solution to a syringe with a tip-blunt capillary (inner diameter = 0.9 mm).

13. Push the solution out of the capillary tip by syringe pump (6 μL/min) and form the ultrathin filament under a strong electric field (1.5 kV DC) on the glass slide in 5 min.

14. Cut and peel off the Scotch tape and leave a 2 mm wide fiber mesh as a barrier near the chamber outlet (Fig. 6b) (*see* **Note 1**).

15. The PDMS housing and the PCL fiber mesh printed glass are treated with oxygen plasma for 1 min and then aligned and bonded to form the integrated device.

3.2 Well Plate-Based Perfusion Culture Device Fabrication

As shown in Fig. 2c, the perfusion culture device consists of: (1) a commercial PS bottomless 96-well plate (Model 655-000, Greiner Bio-One), (2) three PDMS layers with each layer being about 1 mm-thick, (3) PC membrane disks with 6 mm in diameter and 11 μm in thickness, (4) a 1.2 mm-thick glass layer, and (5) PDMS plugs. As illustrated in Fig. 2b, the PDMS layers is used to: (1) provide a fluidic channel of 2 mm thick and 6 mm wide between the inlet and outlet chambers and (2) anchor the placement of the PC membrane disks in the culture chamber during the device assembly.

3.2.1 PDMS Layer 1 Fabrication (Fig. 3)

1. Mix the PDMS precursor and the curing agent at a 10:1 ratio (~30 g total).

2. Pour the mixture on the well plate lid.

3. Degas the mixture in a vacuum desiccator.

4. Place the bottomless 96-well plate onto the PDMS mixture placed on the plate lid while continuing to degas until no bubble can be observed.

5. Cure the assembly in an oven at 70 °C for 5 h.

6. Remove the plate lid from the PDMS/well plate assembly.

3.2.2 PDMS Layer 2 and 3 Fabrication (Fig. 3)

1. Mix the PDMS precursor and the curing agent at a 10:1 ratio (~30 g total).

2. Pour the mixture on the well plate lid.

3. Degas the mixture in a vacuum desiccator.

4. 10 g PDMS precursor mixture are then placed onto the outer side of plate lid and kept horizontal to mold the mixture into the 1 mm-thick flat layer upon curing.

5. Cure the assembly in an oven at 70 °C for 5 h.

6. Peeled off cured PDMS layers from plate lids.
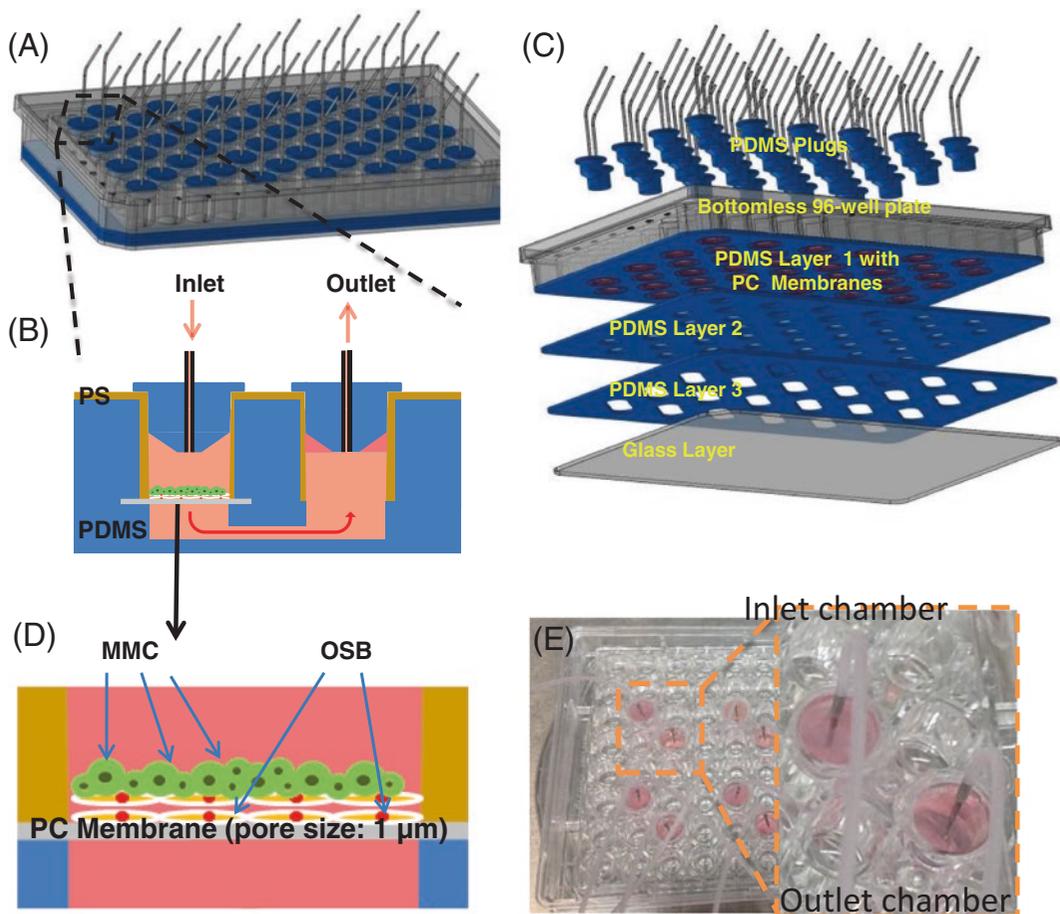
Fig. 2 96-well plate-based perfusion device. (**a–c**) Schematic representations of device design features and fabrication steps. (**d**) Schematic illustration of MMC seeding onto OSB scaffold. (**e**) Actual device, showing polyethylene tubing connected to the stainless steel pins at the inlet and outlet chambers. The metal pins are connected to the PDMS plugs. Reproduced from Zhang W. et al., 2015, DOI: 10.1039/c5lc00341e [42], with permission from The Royal Society of Chemistry Publishing Company

3.2.3 PDMS Plugs Fabrication

1. Mix the PDMS precursor and the curing agent at a 10:1 ratio (~30 g total).
2. Transfer the mixture on the well plate lid.
3. Degas the PDMS mixture in a vacuum desiccator.
4. Place the bottomless 96-well plate onto the PDMS mixture placed on the plate lid without degassing.
5. Cure the assembly in an oven at 70 °C for 5 h.
6. Detach the PDMS structure from bottomless 96-well plate and well plate lid.
7. Cut with a blade to separate column shape PDMS structures as plugs.
8. Punch a hole through the center of plugs using a hand press and a stainless steel round punch (0.6 mm inner diameter).
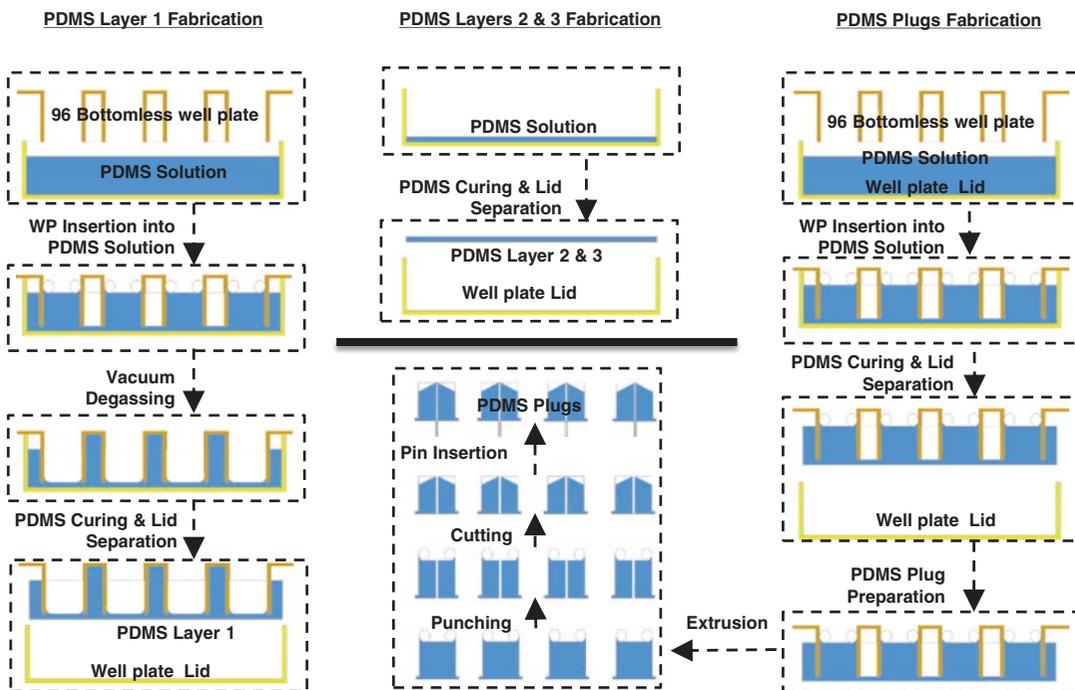
Fig. 3 Schematic illustrations of fabricating PDMS parts for microfluidic connections. Reproduced from Zhang W. et al., 2015, DOI: 10.1039/c5lc00341e [42], with permission from The Royal Society of Chemistry Publishing Company

3.2.4 PDMS Layer Modification (Fig. 4)

1. PDMS Layers 1, 2, and 3 is physically aligned and stacked by hand. A hole puncher (6 mm) is used to punch out holes through the three PDMS layers using a stainless punch die with a diameter of 6 mm (i.e., same as the diameter of wells).

2. In order to make 6 mm wide microfluidic channels between the inlet and the outlet chambers, the channel areas of the PDMS Layer 3 are cut using a razor blade.

3.2.5 Device Assembly (Fig. 5)

1. Separate PDMS layers 1, 2, and 3 by hand.

2. Cut PC membrane disks from PC membrane sheets using a hole puncher (3/8-in.).

3. Place PC membrane disks onto the bottom surface of PDMS Layer 1.

4. Bond PDMS Layer 2 to the bottom surface of PDMS Layer 1 while sandwiching the PC membrane disks using oxygen plasma treatment.

5. Bond PDMS Layer 3 to the bottom surface of PDMS Layer 2 using oxygen plasma treatment.

6. Bond the large glass slide to the bottom surface of PDMS Layer 3 using oxygen plasma treatment.

PDMS Layers Modification
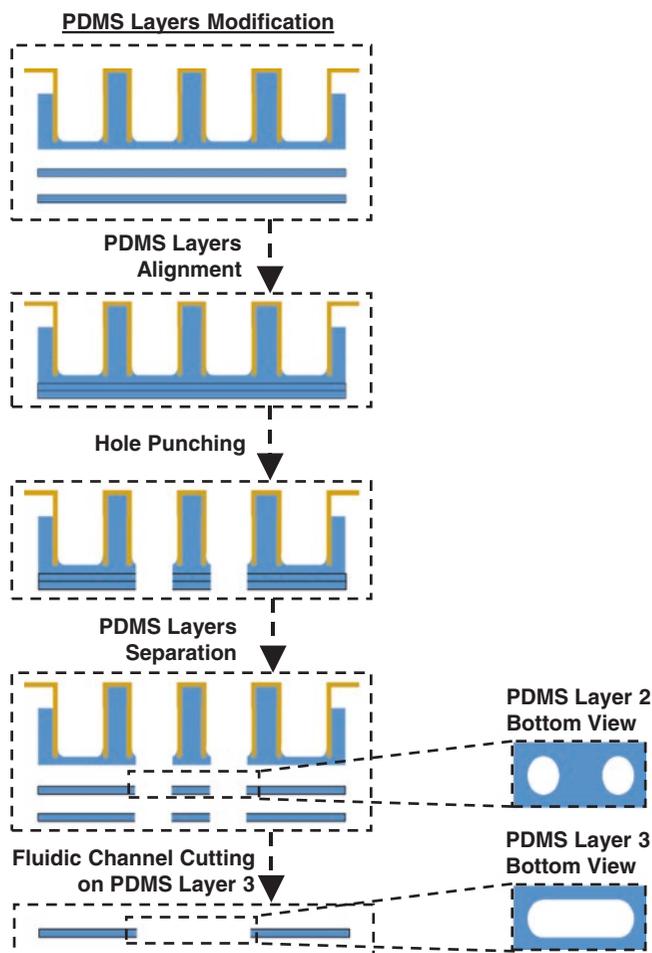
Fig. 4 Schematic illustration of modifying PDMS parts. Reproduced from Zhang W. et al., 2015, DOI: 10.1039/c5lc00341e [42], with permission from The Royal Society of Chemistry Publishing Company

3.3 3D Multiple Myeloma Tissue Model Construction

3.3.1 Ossified Tissue Construction

1. Prepare 0.1 mg/mL human fibronectin in PBS.

2. Maintain hFOB 1.19 cells at 34 °C, 5 % CO_2 incubator until use.

3. Autoclave stainless steel pins and stainless steel dispensing needle prior to culture.

4. Decontaminate each culture compartment. To prepare the microfluidic tissue culture device, pipette 20 μL of 70 % IPA into each culture chamber to sterilize the entire chamber. To prepare the well plate-based perfusion culture device, pipette 300 μL of 70 % IPA into all the inlet and outlet chamber wells to fill and sterilize the entire fluidic passage (*see* **Note 2**).

5. After washing 3× with IPA, rinse the culture compartments of each device 3× with PBS as described in **step 4**.
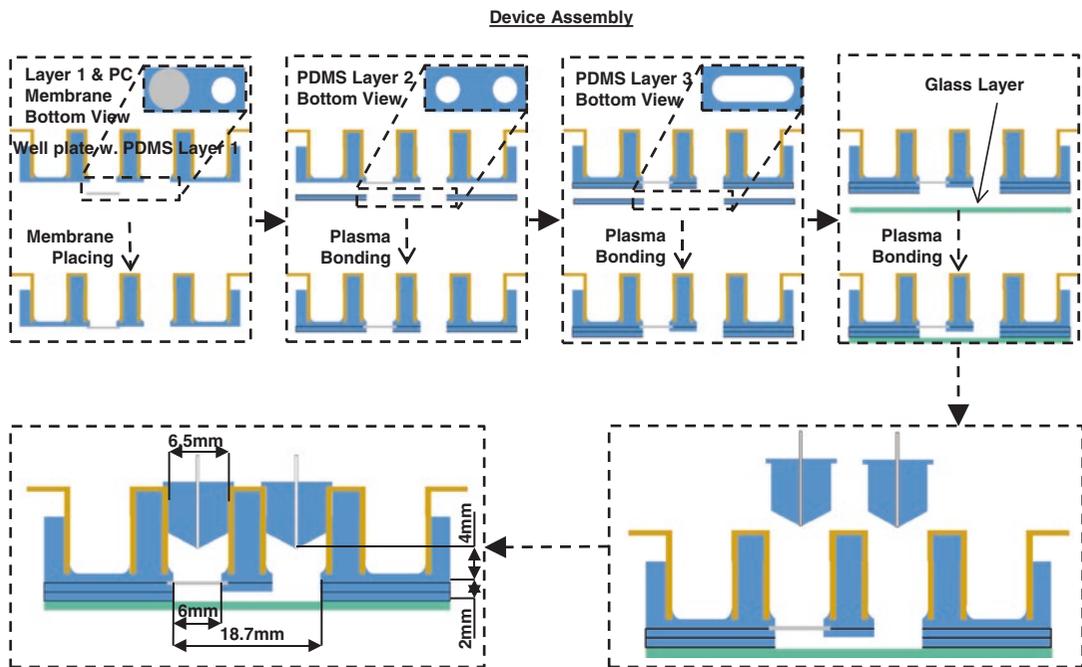
Fig. 5 Schematic illustration of device assembly. Reproduced from Zhang W. et al., 2015, DOI: 10.1039/c5lc00341e [42], with permission from The Royal Society of Chemistry Publishing Company

6. Coat with human fibronectin the culture chambers. To this end, pipette 20 μL of human fibronectin into each chamber of the microfluidic device and 30 μL in each of the inlet chambers of plate-based perfusion culture device. Incubate at 1 h at room temperature.

7. Rinse the fluidic passage of each device 2× with PBS by pipetting in the same manner as in **step 4**.

8. To prepare the devices for cell seeding, replace the PBS by pipetting 20 μL of hFOB 1.19 growth medium into each chamber of the microfluidic tissue culture device. For the well plate-based perfusion culture device, replace the PBS in the inlet chamber with 250 μL of hFOB 1.19 growth medium by pipetting.

9. Seed ~2×10^4 hFOB 1.19 cells into each chamber of the microfluidic tissue culture device by pipetting 10 μL of the cell suspension. To cover the culture wells of the well plate-based perfusion culture device, seed ~2×10^4 hFOB 1.19 cells into each well by pipetting ~50 μL of the cell suspension. Incubate at 34 °C, 5% CO_2 incubator for 2 h.

10. Fill 5 mL of hFOB 1.19 growth medium into each syringe; make sure to remove all the bubbles in the syringe (*see* **Note 3**).

11. Connect syringe with stainless steel dispensing needle (1/2″ Long), which is connected with polyethylene tubing (1 m long).

12. To setup the inlet connections of the culture chambers of the microfluidic tissue culture device, insert 1/3 of a stainless steel pin (0.63 mm outer diameter) into the other side of polyethylene tubing (0.58 mm inner diameter, 1 m long) (Fig. 1a). For the well plate-based perfusion culture device, connect the PDMS plugs and polyethylene tubing (0.58 mm inner diameter, 1 m long) through the stainless steel pins (0.63 mm outer diameter, Fig. 2) (*see* **Note 4**).

13. To setup the outlet connections of the culture chambers of the microfluidic tissue culture device, insert 1/3 of a stainless steel pin (0.63 mm outer diameter) into one side of polyethylene tubing (0.58 mm inner diameter, 20 mm long) (Fig. 1a). In the well-plate perfusion device, the outlet is setup by connecting the PDMS plugs and polyethylene tubing (0.58 mm inner diameter, 20 mm long) through the stainless steel pins (0.63 mm outer diameter) (Fig. 2).

14. Lastly, connect the inlet and outlet connections to each culture chamber of the microfluidic tissue culture device by gently inserting the corresponding stainless steel pins into the inlet and outlet ports in the PDMS (Fig. 1b) using forceps. For the well plate-based perfusion culture device, place the PDMS plugs (Figs. 2c and 3) into corresponding culture chambers using forceps.

15. Connect 8–12 outlet chambers with one VOA vial by inserting the outlet polyethylene tubing through the silicone top of the VOA vial.

16. Place eight syringes in a syringe pump which can be used to control the flow rate of culture medium.

17. Culture hFOB 1.19 cells for 4 days using flow rate of 0.25 μL/min.

3.3.2 Isolation of Bone Marrow Mononuclear Cells (BMMC) Using Ficoll Density Gradient Separation

The bone marrow aspirate is collected in BD Vacutainer® ACD solution A tubes (8.5 mL, Ref# 364606). To isolate BMMC a Fricoll density gradient separation is conducted as follows:

1. Carefully layer the contents of one tube of bone marrow aspirate over 15 mL of Ficoll-Paque Plus (Amersham Biosciences) in a 50 mL conical tube (*see* **Note 5**).

2. Centrifuge at $400 \times g$ for 30–40 min at 20 °C in a swinging-bucket rotor with break off.

3. Aspirate the upper layer and transfer into a new conical tube. This upper layer contains the patient's plasma, which is used to prepare the bone marrow culture medium. The subsequent layer contains the mononuclear cells (lymphocytes, monocytes, and thrombocytes) undisturbed at the interphase.

4. Carefully transfer the mononuclear cell layer to a new 50 mL conical tube.

5. To wash the cells, fill the conical tube with PBS and centrifuge at $300 \times g$ for 10 min at 20 °C (break on). Carefully remove supernatant completely.

Fig. 6a shows the co-localization of CFSE-labeled patient BMMC (green) within the 3D ossified tissue construct (gray background). CFSE-labeling is used, as specified below (Subheading 3.3.4), to quantify cell proliferation. Gross differences between the 3D ossified tissue construct and the tissue construct where BMMC were seeded can be found in Fig 6b. A confocal image of the 3D tissue construct with embedded BMMC after a 21-day coculture period can be found in Fig. 7.
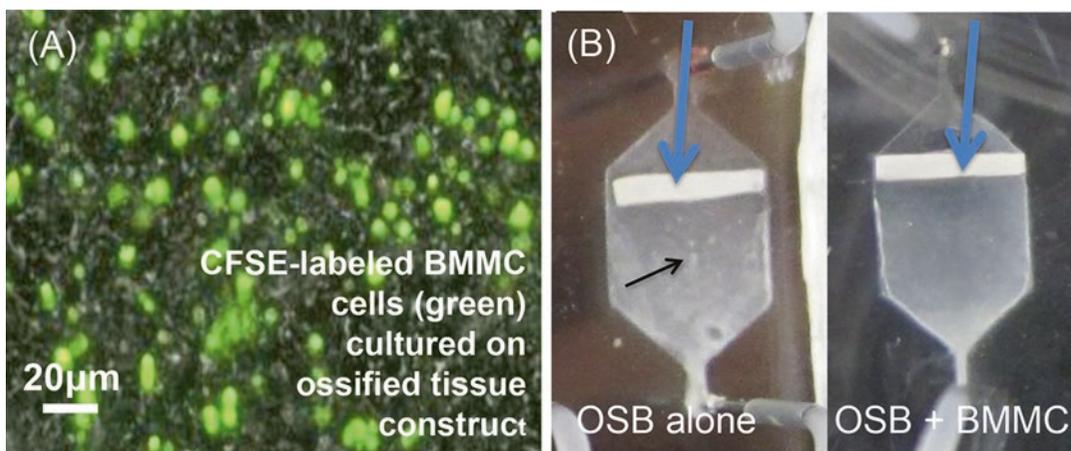
Fig. 6 Representative image of CFSE-labeled BMMC biospecimen from a MM patient cultured in the 3D ossified tissue construct. (**a**) Merged bright flied and fluorescent image of CFSE-labeled BMMC in 3D ossified tissue construct, 8 h after seeding; CFSE+BMMC (*green*), OSB cells and their ECM (*gray/black background*). (**b**) In the absence of patient BMMC, the ossified tissue (day 21) presented with more mineralized particles (*arrow*). BMMC appear to disrupt the deposition of ECM by OSB hence the more homogenous appearance of the ossified tissue (OSB + BMMC). A PCL fiber mesh (*blue arrow*) was placed near the chamber's outlet to ensure retention of non-adherent BMMC within the culture chamber. Reproduced from Zhang W. et al., 2014, DOI: 10.1089/ten.tec.2013.0490 [10], with permission from Mary Ann Liebert, Inc. Publishing Company. (Color figure online)
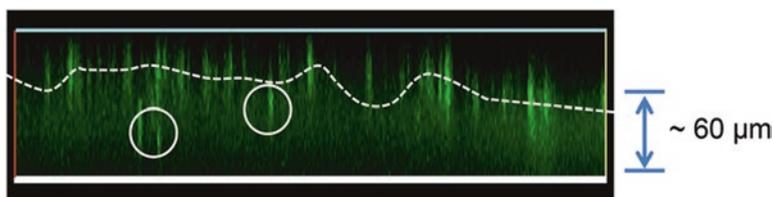
Fig. 7 Confocal image of BMMC in the 3D ossified tissue construct. The stratification of BMMC in the 3D tissue was assessed by confocal microscopy (day 21). Diffused green area represents the ECM thickness of the 3D bone tissue (~60 μm). CFSE+BMMC can be seen as *bright green spots*. Average location of BMMC in the upper surface of the 3D tissue is followed by the *dashed white line*. On day 21 BMMC could be found deeper in the tissue (*white circles*). Reproduced from Zhang W. et al., 2014, DOI: 10.1089/ten.tec.2013.0490 [10], with permission from Mary Ann Liebert, Inc. Publishing Company (Color figure online)

3.3.3 BMMC Cryopreservation **(See Note 6***)*

1. BMMC are counted in a hemocytometer and reconstituted at $0.5–1 \times 10^7$ cell/mL in 10% dimethyl sulfoxide (DMSO, biotechnology performance certify, Sigma-Aldrich, D2438) and 90% FBS.

2. The cell suspension is then aliquoted into cryovials (CryoElite, yellow caps, Wheton) (1 mL/vial).

3. Vials can be frozen and stored in liquid nitrogen until use.

3.3.4 Seeding of Patient BMMC

1. Take one vial of patient BMMC sample out of liquid nitrogen. Thaw patient's BMMC by placing the frozen cryogenic vials in 37 °C water bath for about 1 min until ~90% of the sample is thawed.

2. Transfer the entire sample into a 50 mL conical tube containing 30 mL of cRPMI medium with 1 unit/mL heparin (*see* **Note 7**) and spin down the BMMC.

3. Remove the supernatant (in cRPMI medium) and resuspend BMMC by adding 1 mL complete bone marrow culture medium.

4. Filter the BMMC suspension through 40 μm cell strainer, and collect the filtered BMMC suspension in a 1.5 mL microcentrifuge tube.

5. Split 1 mL BMMC suspension into four wells of a 96-well plate and incubate in 37 °C, 5% CO_2 incubator for 1 h.

6. To identify the percentage of MM cells within the patient BMMC before culture, take 50 μL a BMMC suspension and split into five wells of a 96-well plate for antibody staining (in FACS buffer) per manufacturer's instructions (*see* **Note 8**): (1) anti-CD38 and anti-CD56, (2) anti-CD38 and anti-CD138, (3) anti-CD56 and anti-CD38, and (4) 7-AAD.

7. Acquire a minimum of 10,000 events and analyze by flow cytometry. Identify MM subpopulations by double gating on CD38+CD56+, CD38+CD138+, and CD56+CD138+ cells.

8. For proliferation assays, label BMMC samples with 0.5 μM CFSE per manufacturer's instruction.

9. Replace hFOB growth medium with 200 μL bone marrow culture medium in the inlets of culture chambers.

10. Seed 4 to 8×10^4 CFSE-labeled BMMC into the 4 day old ossified tissue culture chambers.

11. Stop the flow to allow the settling of cells and reinstate after 4 h with the bone marrow culture medium. Perfusion cultures are kept in a humidified incubator at 37 °C in 5% CO_2 until the termination of the experiments.

12. The MM tumor model can be cultured for up to 4 weeks (*see* **Note 9**).

3.3.5 Characterization of Cell Populations Using Flow Cytometry

1. At the conclusion of the culture, cells can be removed from culture by trypsinization. To this end, add 50 μL trypsin–EDTA, incubate in 37 °C incubator for 10 min. Add 100 μL bone marrow culture medium into each culture well to stop the reaction. Pipet the medium in the well up and down five times before aspirate the medium out with tissue. Transfer the harvested tissue in a microcentrifuge.

2. Further dissociated the harvested tissue by grinding in microcentrifuge tubes using a pellet pestle.

3. Prepared single cell suspensions by filtering the ground tissue through a 40 μm cell strainer.

4. Divide the harvested single cell suspensions from the same culture well into five different wells of a 96-well plate.

5. Stain cells in FACS buffer with the following fluorochrome-conjugated mAb per manufacturer's instructions: (1) anti-CD13, anti-CD38, and anti-CD56, (2) anti-CD13, anti-CD38, and anti-CD138, (3) anti-CD13, anti-CD56, and anti-CD38, and (4) 7-AAD.

6. Acquire a minimum of 10,000 events and analyze by flow cytometry. Triple gate on CFSE$^+$CD38$^+$CD56$^+$ cells, CFSE$^+$CD38$^+$CD138$^+$ and CFSE$^+$CD56$^+$CD138$^+$ cells to identify MM subpopulations (Fig. 8). CFSE$^-$CD13$^+$7-AAD$^+$ and CFSE$^+$ 7-AAD$^+$ cells are discerned as dead OSB and dead BMMC respectively.

7. Determine the MMC population expansion and percentage in BMMC (*see* **Note 10**).

4   Notes

1. A PCL fiber mesh can be placed near the outlet of the culture chamber by electrospinning [43] in order to retain non-adherent BMMC, which would otherwise end up flowing through the system due to the direction of the flow in the microfluidic tissue culture system (i.e., side to side). Of note, this is not an issue in the 96-well perfusion culture device because the flow direction in this culture platform is from top to bottom and the PC membrane in the inlet well acts also as a retainer barrier preventing non-adherent BMMC to flow out of the culture chamber.

2. There are two small holes (0.6 mm in diameter) in each chamber of microfluidic tissue culture device for the inlet/outlet connections (i.e., inlet and outlet ports, Fig. 1b). In order to pipette liquid into the chamber, through the ports, a 1–100 μL pipette tip has to be used to match the port size for liquid injection as shown in Fig. 9.
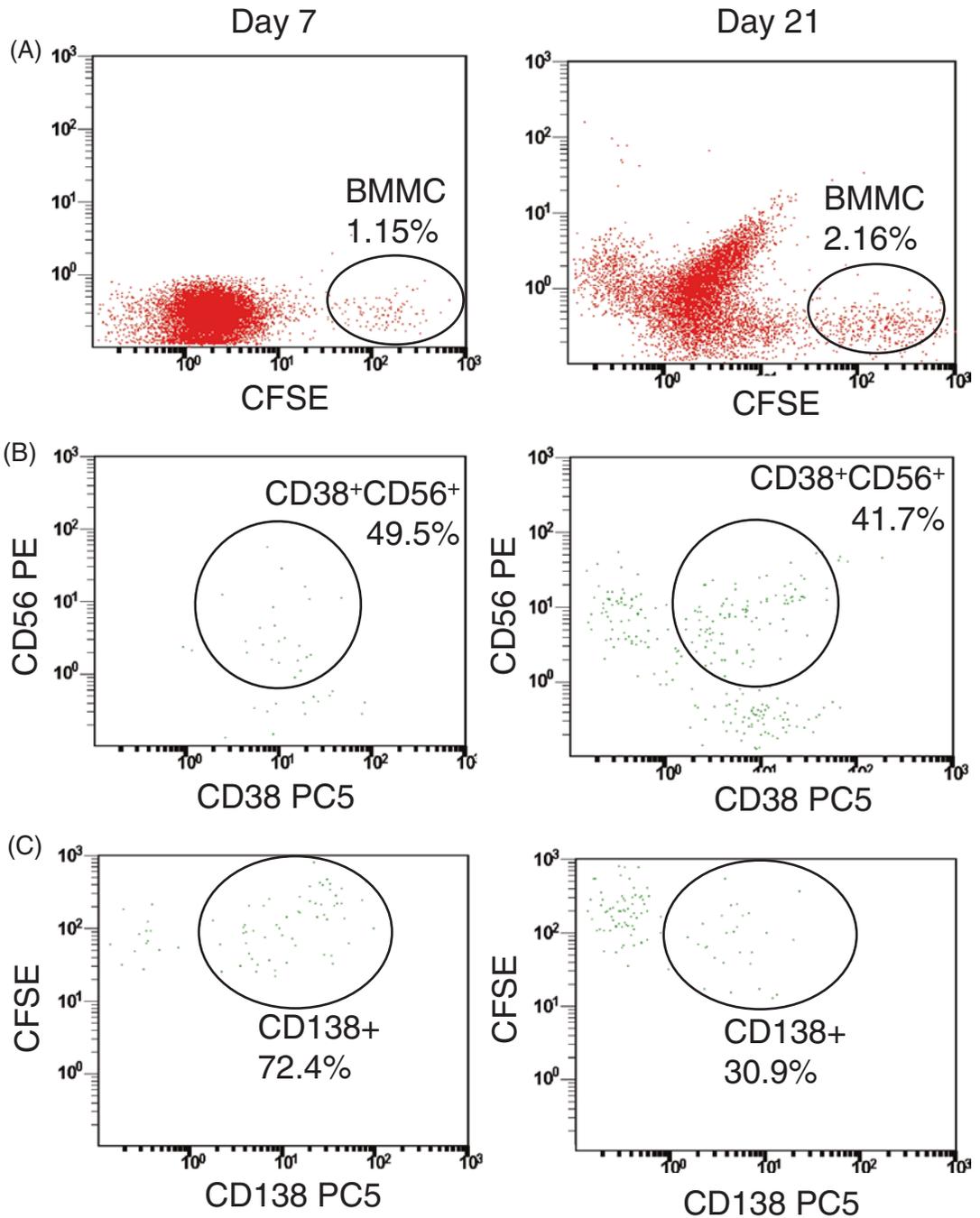
Fig. 8 Representative dot plots images from flow cytometric analyses of a patient on days 7 and 21 of culture. (**a**) CFSE+BMMC. (**b**) CD38+CD56+ MMC, and (**c**) CD138+ MMC. MMC were gated on the CFSE+BMMC population. Reproduced from Zhang W. et al., 2014, DOI: 10.1089/ten.tec.2013.0490 [10], with permission from a Mary Ann Liebert, Inc. Publishing Company
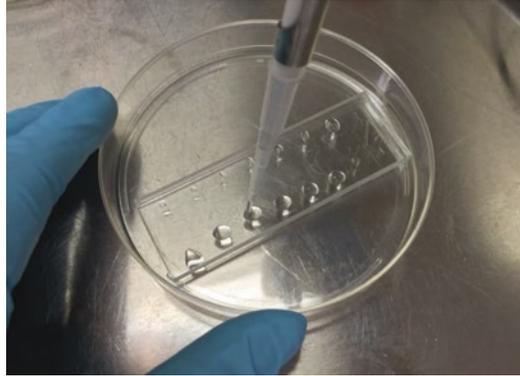
Fig. 9 Pipetting into the microfluidic tissue culture device. A 1–100 µL pipette tip has to be used in order to match the hole for liquid injection into the culture chambers

3. After aspirating hFOB growth medium into the syringe, place the dispensing needle (1–1/2″ Long) onto syringe. Push syringe to let medium fill up the needle, then use forceps to gently tap the syringe to let bubble rise up to the top of the syringe. Push the bubble out of the needle.

4. Connect stainless steel pins with polyethylene tubing first; let the polyethylene tubing covers 1/3 of the stainless steel pins. Let the other 2/3 of a stainless steel pin go through a PDMS plug.

5. Bone marrow aspirates can be diluted with three volumes of PBS before layering on Ficoll in order to increase the yield of BMMC and decrease the amount of red blood cells.

6. Our culture experiments were conducted using frozen BMMC samples. For all samples evaluated in our studies (>20 to date), we have successfully been able maintain and expand every MMC subpopulation tested, therefore we anticipate the same success rate when using freshly isolated (i.e., not cryopreserved) BMMC.

7. Upon thawing the BMMC it is common to find a gelatinous-like aggregate. Heparin is added to dissolve this type of aggregates, which interferes with counting and plating of cells.

8. While antibodies can be used at the concentrations indicated by the manufacturer, we have found that a lower concentration is equally efficacious for labeling the desired cell populations:

    (a) Staining of cell is conducted in a 96 round bottom-well plate. Typically, after culture, dissociation and reconstitution, we place around 1×10^4 to 2×10^4 cells/well in ~5 µL FACS buffer in each well for staining.

    (b) We stained with 1 µL of antibody per well although recommended standard is higher for Beckman Coulter and BD Pharmingen antibodies.

(c) For 7-AAD staining, bring the volume of the well to 250 μL with FACS buffer, add 5 μL of 7-AAD and incubate for 10 min right before test to measure the percentage of dead cells by flow cytometry.

9. To support the long-term culture (>3 weeks) of MMC, 1×10^4 to 2×10^4 healthy OSB can be reintroduced in the 96-well plate perfusion culture device (by simply removing the inlet plugs and adding the cells) after a 2-week culture period to bypass the deleterious effects of MMC on OSB activity and viability. This strategy can significantly extend the survival of viable MMC pass 4 weeks culture [9].

10. CFSE intensity is used to determine MMC proliferation, as each cell division results in a 50% decrease in the florescence intensity of this dye. Division peaks are labeled from 0 to n. Since a single cell dividing n times will generate $2n$ daughter cells, and, if the total number of cells resulting from three divisions ($n = 3$) is eight, then exactly one mother cell had to divide three times to generate these eight daughter cells ($2^3 = 8$). Making use of this mathematical relationship, the number of divisions of mother cells was extrapolated from the number of daughters under each division peak. The average cell expansion is then calculated as described [10]:

$$\text{Average cell expansion} = \frac{Y}{M} \tag{1}$$

Where:

$$Y = \text{yield cell number}, \ M = \text{mother cell number} \tag{2}$$

$$M = \sum_n Y * \frac{\% \text{ of cells under division peak}}{2^n} \tag{3}$$

Therefore:

$$\text{Average cell expansion} = \frac{1}{\sum_n \left(\% \text{ of cells under division peak}\right) / 2^n} \tag{4}$$

5 Acknowledgements

References

1. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical development success rates for investigational drugs. Nat Biotechnol 32:40–51

2. Cook D, Brown D, Alexander R et al (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat Rev Drug Discov 13:419–431

3. Li A, Walling J, Kotliarov Y et al (2008) Genomic changes and gene expression profiles reveal that established glioma cell lines are poorly representative of primary human gliomas. Mol Cancer Res 6:21–30

4. Domcke S, Sinha R, Levine DA et al (2013) Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun 4:2126

5. Sandberg R, Ernberg I (2005) Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). Proc Natl Acad Sci U S A 102:2052–2057

6. Aggeler J, Park CS, Bissell MJ (1988) Regulation of milk protein and basement membrane gene expression: the influence of the extracellular matrix. J Dairy Sci 71:2830–2842

7. Bhat R, Bissell MJ (2014) Of plasticity and specificity: dialectics of the micro- and macro-environment and the organ phenotype. Wiley Interdiscip Rev Membr Transp Signal 3:147–163

8. Bissell MJ, Hall HG, Parry G (1982) How does the extracellular matrix direct gene expression? J Theor Biol 99:31–68

9. Zhang W, Gu Y, Sun Q et al (2015) Ex vivo maintenance of primary human multiple myeloma cells through the optimization of the osteoblastic niche. PLoS One 10:e0125995. doi:10.1371/journal.pone.0125995

10. Zhang W, Lee WY, Siegel DS et al (2014) Patient-specific 3D microfluidic tissue model for multiple myeloma. Tissue Eng Part C Methods 20:663–670

11. Sachs N, Clevers H (2014) Organoid cultures for the analysis of cancer phenotypes. Curr Opin Genet Dev 24:68–73

12. Buske P, Przybilla J, Loeffler M et al (2012) On the biomechanics of stem cell niche formation in the gut—modelling growing organoids. FEBS J 279:3475–3487

13. Gu Y, Zhang W, Qiaoling S et al (2015) Microbead-guided reconstruction of the 3D osteocyte network during microfluidic perfusion culture. J Mater Chem B 3:3625–3633

14. Damiano JS (2002) Integrins as novel drug targets for overcoming innate drug resistance. Curr Cancer Drug Targets 2:37–43

15. Li ZW, Dalton WS (2006) Tumor microenvironment and drug resistance in hematologic malignancies. Blood Rev 20:333–342

16. Nakagawa Y, Nakayama H, Nagata M et al (2014) Overexpression of fibronectin confers cell adhesion-mediated drug resistance (CAM-DR) against 5-FU in oral squamous cell carcinoma cells. Int J Oncol 44:1376–1384

17. Schmidmaier R, Baumann P (2008) ANTI-ADHESION evolves to a promising therapeutic concept in oncology. Curr Med Chem 15:978–990

18. Tolias P, Toruner GA (2014) Personalized medicine. Future Med 7:461–464

19. Bhatia SN, Ingber DE (2014) Microfluidic organs-on-chips. Nat Biotechnol 32:760–772

20. Benien P, Swami A (2014) 3D tumor models: history, advances and future perspectives. Future Oncol 10:1311–1327

21. Wang C, Tang Z, Zhao Y et al (2014) Three-dimensional in vitro cancer models: a short review. Biofabrication 6:022001

22. Hickman JA, Graeser R, de Hoogt R et al (2014) Three-dimensional models of cancer for pharmacology and cancer cell biology: capturing tumor complexity in vitro/ex vivo. Biotechnol J 9:1115–1128

23. Birmingham E, Kreipke TC, Dolan EB et al (2015) Mechanical stimulation of bone marrow in situ induces bone formation in trabecular explants. Ann Biomed Eng 43:1036–1050

24. Jaasma MJ, Plunkett NA, O'Brien FJ (2008) Design and validation of a dynamic flow perfusion bioreactor for use with compliant tissue engineering scaffolds. J Biotechnol 133:490–496

25. Plunkett N, O'Brien FJ (2011) Bioreactors in tissue engineering. Technol Health Care 19:55–69

26. Ferrarini M, Steimberg N, Ponzoni M et al (2013) Ex-vivo dynamic 3-D culture of human tissues in the RCCS bioreactor allows the study of multiple myeloma biology and response to therapy. PLoS One 8:e71613

27. Kwapiszewska K, Michalczuk A, Rybka M et al (2014) A microfluidic-based platform for tumour spheroid culture, monitoring and drug screening. Lab Chip 14:2096–2104

28. Lee JH, Gu Y, Wang H, Lee WY (2012) Microfluidic 3D bone tissue model for high-throughput evaluation of wound-healing and infection-preventing biomaterials. Biomaterials 33:999–1006

29. Lei KF, Wu MH, Hsu CW, Chen YD (2014) Real-time and non-invasive impedimetric monitoring of cell proliferation and chemosensitiv-

ity in a perfusion 3D cell culture microfluidic chip. Biosens Bioelectron 51:16–21

30. Polini A, Prodanov L, Bhise NS et al (2014) Organs-on-a-chip: a new tool for drug discovery. Expert Opin Drug Discovery 9:335–352

31. Ocio EM, Richardson PG, Rajkumar SV et al (2014) New drugs and novel mechanisms of action in multiple myeloma in 2013: a report from the International Myeloma Working Group (IMWG). Leukemia 28:525–542

32. Lonial S, Anderson KC (2014) Association of response endpoints with survival outcomes in multiple myeloma. Leukemia 28:258–268

33. Neri P, Bahlis NJ (2012) Targeting of adhesion molecules as a therapeutic strategy in multiple myeloma. Curr Cancer Drug Targets 12:776–796

34. Neri P, Ren L, Azab AK et al (2011) Integrin beta7-mediated regulation of multiple myeloma cell adhesion, migration, and invasion. Blood 117:6202–6213

35. Damiano JS, Dalton WS (2000) Integrin-mediated drug resistance in multiple myeloma. Leuk Lymphoma 38:71–81

36. Abdi J, Chen G, Chang H (2013) Drug resistance in multiple myeloma: latest findings and new concepts on molecular mechanisms. Oncotarget 4:2186–2207

37. Yaccoby S, Epstein J (1999) The proliferative potential of myeloma plasma cells manifest in the SCID-hu host. Blood 94:3576–3582

38. Yaccoby S, Barlogie B, Epstein J (1998) Primary myeloma cells growing in SCID-hu mice: a model for studying the biology and treatment of myeloma and its manifestations. Blood 92:2908–2913

39. Lawson MA, Paton-Hough JM, Evans HR et al (2015) NOD/SCID-GAMMA mice are an ideal strain to assess the efficacy of therapeutic agents used in the treatment of myeloma bone disease. PLoS One 10:e0119546

40. Reagan MR, Mishima Y, Glavey SV et al (2014) Investigating osteogenic differentiation in multiple myeloma using a novel 3D bone marrow niche model. Blood 124:3250–3259

41. Kirshner J, Thulien KJ, Martin LD et al (2008) A unique three-dimensional model for evaluating the impact of therapy on multiple myeloma. Blood 112:2935–2945

42. Zhang W, Gu Y, Hao Y et al (2015) Well plate-based perfusion culture device for tissue and tumor microenvironment replication. Lab Chip 15:2854–2863

43. Yang X, Ogbolu KR, Wang H (2008) Multifunctional nanofibrous scaffold for tissue engineering. J Exp Nanosci 3:329–345

Chapter 13

microRNA Target Prediction

William Ritchie

Abstract

microRNAs are short RNAs that reduce gene expression by binding to their targets. Computational predictions indicate that all human genes may be regulated by microRNAs, with each microRNA possibly targeting thousands of genes. Commonly used software will produce a prohibitive number of predicted targets for each microRNA. Here I describe procedures that refine these predictions by integrating available software and expression data from experiments available online. These procedures are tailored to experiments where predicting true targets is more important than detecting all putative targets.

Key words microRNA, Target genes, microRNA expression

1 Introduction

microRNAs (miRNAs) are short, ~22 nucleotide long RNAs that reduce gene expression, usually by binding to the 3' untranslated region of target mRNAs. miRNAs guide a protein complex called RNA-induced silencing complex (RISC) to specific mRNA target sites. miRNAs were first discovered in 1993 during an analysis of larval developmental timing in the worm *Caenorhabditis elegans*, where a 22-nucleotide RNA regulated protein abundance of LIN-14 [1]. First regarded as a Nematode-specific RNA family, it was only in 2000 that another microRNA, let-7, was characterized and identified in other species. In 2002, Eric Lai compared the sequences of 11 microRNAs to the K box and Brd Box motifs that were known to mediate post-transcriptional regulation in Drosophila. He demonstrated that the first eight nucleotides, now called the seed region, of miRNAs, were perfectly complementary to these motifs and concluded that this complementarity may be essential in posttranscriptional regulation by microRNAs [2]. This simple bioinformatics analysis established one of the strongest predictive features used in target prediction to date and was the basis for numerous algorithms that enabled the explosion of miRNA functional characterization.

To date over 1000 miRNAs have been identified in humans, hundreds of which are associated with major biological processes including cell proliferation and differentiation, development, and disease. As such miRNAs are arguably one of the most important classes of functional RNAs. However, the rules governing miRNA target recognition are not fully understood and may vary for each miRNA–mRNA pair. Computational approaches that can test various models of miRNA binding and predict target sites are therefore essential to understanding the function of microRNAs.

Here I present three strategies that decrease the number of false positive predicted targets. In some cases this comes at the cost of more false negatives. The first technique called multi-targeting takes advantage of the fact that microRNAs that target the same gene multiple times can be detected with higher signal to noise ratios than those that target the same gene once [3]. The second technique uses readily available expression data to identify mRNAs for which the expression is dependent on changes in microRNA expression. The expression of microRNAs that inhibit mRNAs should be negatively correlated with the expression of their targets. The third integrates databases that collect experimental evidence for specific targets.

In this guide, I use the TargetScan [4] algorithm and the miRBase version of the miRanda [5] algorithm. I selected these two algorithms firstly because they displayed excellent performance in terms of sensitivity and specificity in a recent thorough benchmark [6]. Secondly, these algorithms use different criteria to predict targets and thus I consider them to be complementary and results from both should be considered for subsequent investigation. The principles underlying these two algorithms are explained below (*see* **Notes 1–4**). I also use online databases that compile lists of experimental evidence supporting microRNA targets. This evidence ranges from the "gold standard" Luciferase assay to much weaker evidence such as changes in mRNA expression subsequent to microRNA knockdown or knockin. Finally, I use the mimiRNA [7] server that provides useful tools for the analysis of microRNA targets such as expression profiling and multi-targeting.

2   Materials

1. A computer with an Internet connection.
2. Microsoft Excel or its equivalent in free software such as OpenOffice or StarOffice.

3   Methods

3.1   Using Multi-targeting to Discover microRNA Targets (Miranda and TargetScan)

3.1.1   With TargetScan

Recent analysis demonstrates that numerous mRNAs are targeted by the same miRNA at different sites within their 3' UTR [3] (Fig. 1, bottom panel). This multi-targeting occurs at a significantly higher rate than expected. Focusing therefore on mRNAs that have more than one predicted site for the same miRNA in the 3' UTR can increase the signal to noise ratio for different algorithms [3]. Although this approach will eliminate numerous true target sites it has the advantage of producing a list of high confidence gene targets. This method requires the user to first select one or more target prediction programs and subsequently refine their results for multi-targeting. In a similar approach, the PicTar [8] algorithm uses a combinatorial approach that not only accounts for multiple binding sites of the same miRNA but also computes the likelihood that a sequence is bound by a combination of input miRNA sequences (Fig. 1, top panel).

3.1.2   How to: Filtering Predictions with Multi-targeting

1. Go to TargetScan (http://www.targetscan.org/). Select the species, paste or type your microRNA identifier in the section "Enter a microRNA name," and click "Submit."

2. In the results page, in the "Conserved sites" column, in the sub-column named "total," search for the number 3 or above. This corresponds to genes with at least three conserved predicted sites for the inputted microRNA. The gene symbols corresponding to these rows in the left column are the targets of interest. If no such genes are found, the threshold should not be reduced to select genes with less sites because there is insufficient proof that this will increase prediction accuracy. Users should instead use the same approach with the miRanda algorithm or use the approach described in Subheading 3.2.
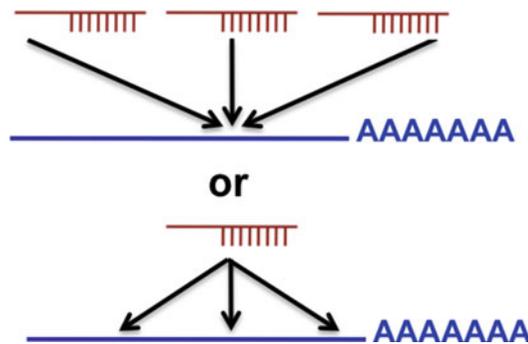
Fig. 1 Different miRNAs that target the same gene (*top panel*), and an miRNA that targets the same gene at different locations can reduce false positive predictions (*bottom panel*)

3.1.3   With miRanda

1. Go to the miRBase implementation of miRanda (http://www.ebi.ac.uk/enright-srv/microcosm). Click on the Search button.

2. Select the genome, paste your microRNA identifier in the section "Enter microRNA id," and click "Search."

3. At the top of the Target Listing page, next to "Download Table," click on the "TXT" link. This allows you to download the results in text format. Your browser should prompt you to save this file. Save it in a convenient place on your computer.

4. Right click (Ctrl + click on Mac) the saved folder and select "Open With," then Excel. Make sure Excel recognizes the individual columns (there should be 13 columns) of this file. Select the column with gene names (column "M"). Copy this column.

5. Go to mimiRNA's multi-tool (http://mimiRNA.centenary.org.au/mep/MultiT.htm) and paste the list in the text area next to "Input list of microRNAs or Genes." Click "Find multiple occurrences." The results page will show you genes with 3 or more targets for your microRNA. This function of mimiRNA can be used for lists of genes or microRNAs from any source.

3.2   Using Expression Data to Discover Targets

Numerous miRNAs inhibit gene expression by destabilizing mRNAs [9]. As a consequence, mRNA targets should be expressed at lower levels in tissues where the miRNA is expressed. Correlating mRNA and miRNA expression across multiple tissues and selecting those pairs that are negatively correlated can successfully detect target genes (Fig. 2) [10]. Because this method is independent of any sequence analysis, it can be used to filter predictions made by any of the aforementioned algorithms. Another advantage of this approach is that it is not restricted to targets located in the 3' UTR. Although there are fewer published examples of miRNA targets in other regions of mature mRNAs, there may be numerous targets in the coding region that have been overlooked because the high level of sequence conservation in exons prohibits the use of sequence conservation-based techniques. The major drawback of this approach is that miRNAs that do not affect mRNA levels or that only "fine-tune" gene expression will not be identified. The mimiRNA website [7] provides correlation analysis in human samples and displays the predicted targets from TargetScan, miRanda, and PicTar.

How to: Integrating Expression Data

1. Go to mimiRNA (http://mimiRNA.centenary.org.au). Click option 5 "Which genes does my microRNA target."

2. Select your microRNA in the scroll down menu. In the "Integrate with data from" section, select "targetScan" or "miRanda". This will allow you to integrate predictions of these algorithms with those made by mimiRNA. Click "Find correlated genes."
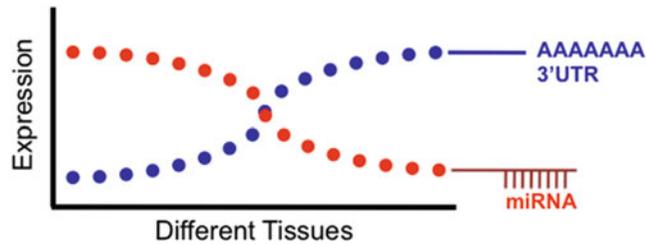
Fig. 2 Negative correlations between an miRNA and its target can help identify biologically relevant targets

3. The right column is a list of gene symbols for which the expression is negatively correlated with the input microRNA and are therefore potential targets. Genes symbols with a ">>>" symbol were also predicted targets according to TargetScan (or miRanda). Each gene can be clicked to go to its Entrez Gene description at the NCBI website.

3.3   Integrating Data from miRNA Target Databases

Recent benchmarking of target prediction algorithms showed that a large portion of confirmed miRNA–target interactions could not be identified even by the most sensitive approaches [11]. Other studies have demonstrated that many algorithms produce conflicting results not only because of the differences in how miRNA–mRNA targets are modeled but also because different online algorithms use different reference gene and 3' UTR databases [3]. For these reasons, curated repositories of experimentally verified targets have become necessary to evaluate existing algorithms but also to develop novel approaches to detect targets. The curation of these databases is tedious because numerous experimental techniques can be used to validate targets and each of these techniques has its own variants. The luciferase reporter assay for example can be used to demonstrate repression of a target in a specific genomic and cellular context but rigorous controls are essential. These controls include whether mutagenesis of the binding site is performed to identify the seed region, whether the 3' UTR context of the original target was conserved or whether the cellular context is similar to the in vivo interaction. Moreover, publishers do not require that experimentally validated miRNA targets be uploaded to a common online resource prior to publication. Each database must therefore extract these data from articles published online. Here I give a brief description of some of the most popular databases.

3.3.1   mirRecords

The first version of miRecords [12] was released in 2008 and was last updated in April 2013. miRecords contains manually curated experimental evidence for 2705 records of interactions between 644 miRNAs and 1901 target genes in nine animal species. It also hosts predicted miRNA target results produced by 11 miRNA target

prediction programs (including all algorithms discussed in this chapter). The database provides information regarding the type of experimental validation performed, the PubMed ID, and a short excerpt from the original article describing the validation approach. Results cannot be filtered on a specific type of experimental evidence.

3.3.2 StarBase

StarBase [13] version 2.0 was released in 2013. This database takes advantage of recent high-throughput experimental technologies such as CLIP-Seq (HITS-CLIP and PAR-CLIP) and degradome sequencing across six organisms. StarBase integrates 108 datasets from 37 different studies with miRNA–mRNA target prediction from six different algorithms to produce ~500,000 targets. StarBase also provides a Genome Browser and functional annotation tools linked to five different gene ontologies to discover enriched biological functions or pathways associated with their targets.

3.3.3 TarBase

Tarbase [14], first introduced in 2005, is currently in its sixth version. TarBase hosts a collection of over 65,000 manually curated experimentally validated miRNA–gene interactions. TarBase integrates data from high-throughput techniques as well as individual miRNA studies. These studies provide either direct or indirect evidence of miRNA–mRNA pairs. The user can filter results based on the type of evidence and the experimental approach used. Users must sign up to use the software and an e-mail request is necessary to obtain the entire dataset.

How to: Combining Experimental Evidence from TarBase

1. Go to mimiRNA's experimental evidence webtool (http:// mimiRNA.centenary.org.au/mep/expEv.htm). This tool combines an input list of targets with experimental evidence taken from the TarBase.

2. Select an organism and a microRNA of interest from the pull-down menus. If the microRNA of interest is not in the pull-down menu, this means that there is no experimental evidence supporting it from TarBase. In the text box next to "Input list of microRNA targets," paste a list of microRNA targets, each target separated by a new line. The input can be a list of targets found by miRanda as described in Subheading 3.1.3 or by any other target prediction program.

3. "Combine experimental evidence." This will display the common targets between the input list and targets with experimental evidence compiled in TarBase.

   Investigators can further explore the type of evidence supporting targets by selecting the microRNA, its target gene and organism at http://diana.cslab.ece.ntua.gr/tarbase/. Clicking on the "+" sign next to "Experimental Conditions" on the results page will show if the evidence is strong (existence of Direct Support) or weak (Indirect Support).

4   ᵓNotes

More information on TargetScan and Miranda:

1. There are many available programs dedicated to microRNA target prediction. In this chapter, I use TargetScan and the miRBase version of miRanda for three main reasons. They have relatively low false-positive rates (<30 %) and good overall performance [6], they are hosted on reliable servers but also the two approaches (and therefore the set of targets they predict) are different.

2. The miRBase version of miRanda searches for exact matches between the 3ᵓ UTR of mRNA genes and the seed region (generally nucleotide positions 2–9) of the microRNA [15]. It then uses a thermodynamic model to evaluate whether the duplex formed between the whole microRNA sequence and the mRNA gene is stable. The microRNA–mRNA pairs for which these conditions are true in at least two other species are retained as predictions.

3. TargetScan v5 relies on the complementarity between the mRNA and the microRNA seed sequences. Each seed that binds to an mRNA sequence will be attributed a score based on the degree of conservation of the region to which the seed binds, whether the complementarity to the seed region is 8 or 7 nucleotides long and whether it has an A at position 1. TargetScan also calculates a context score, which depends on the AU composition of the flanking region and the distance of the target site to the 3ᵓ UTR end.

4. TargetScan is therefore biased towards conserved microRNA targets whereas miRanda is biased towards microRNAs that bind with higher calculated affinity to their target.

References

1. Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854

2. Lai EC (2002) Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet 30:363–364. doi:10.1038/ng865

3. Ritchie W, Flamant S, Rasko JE (2009) Predicting microRNA targets and functions: traps for the unwary. Nat Methods 6:397–398. doi:10.1038/nmeth0609-397, nmeth0609-397 [pii]

4. Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19:92–105. doi:10.1101/gr.082701.108, gr.082701.108 [pii]

5. John B, Enright AJ, Aravin A et al (2004) Human MicroRNA targets. PLoS Biol 2:e363

6. Fan X, Kurgan L (2014) Comprehensive overview and assessment of computational prediction of microRNA targets in animals. Brief Bioinform 16:780–794. doi:10.1093/bib/bbu044

7. Ritchie W, Flamant S, Rasko JE (2010) mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. Bioinformatics 26:223–227. doi:10.1093/bioinformatics/btp649, btp649 [pii]

8. Krek A, Grün D, Poy MN et al (2005) Combinatorial microRNA target predictions. Nat Genet 37:495–500

9. Lim LP, Lau NC, Garrett-Engele P et al (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433:769–773

10. Ritchie W, Rajasekhar M, Flamant S, Rasko JE (2009) Conserved expression patterns predict microRNA targets. PLoS Comput Biol 5:e1000513. doi:10.1371/journal.pcbi.1000513

11. Alexiou P, Maragkakis M, Papadopoulos GL et al (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. Bioinformatics 25:3049–3055. doi:10.1093/bioinformatics/btp565

12. Xiao F, Zuo Z, Cai G et al (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 37:D105–D110. doi:10.1093/nar/gkn851

13. Yang JH, Li JH, Shao P et al (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res 39:D202–D209. doi:10.1093/nar/gkq1056

14. Vergoulis T, Vlachos IS, Alexiou P et al (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res 40:D222–D229. doi:10.1093/nar/gkr1161

15. Lewis BP, Shih IH, Jones-Rhoades MW et al (2003) Prediction of mammalian microRNA targets. Cell 115:787–798, doi:S0092867403010183 [pii]

Chapter 14

Evaluating the Delivery of Proteins to the Cytosol of Mammalian Cells

Andrea L.J. Marschall, Congcong Zhang, and Stefan Dübel

Abstract

Delivery of proteins to the cytosol of living cells is a promising research tool. Delivery of antibodies in particular bears exciting applications such as in vivo tracking of proteins at endogenous expression levels or interference with cellular processes. In spite of the large number of methods published for protein delivery, successful applications so far are rare. A possible explanation for this is a vast overestimation of the delivery efficiency due to the use of inappropriate detection methods and/or unsuitable positive controls for cytosolic delivery. Therefore, we provide strategies for unequivocally detecting cytoplasmic protein delivery and quantifying protein transformation efficiency. Finally, we present a protocol for efficient protein delivery to the cytosol validated using these methods.

Key words Antibody, Profection, Protein transfection, Protein delivery, Cell-penetrating peptide (CPP), Protein transduction domain (PTD), Electroporation, Yumab

1 Introduction

A large number of reports are published on the delivery of proteins to the cytosol of cells. The delivery of antibodies in particular to the cytosol or the nucleus could lead to exciting new applications such as in vivo imaging of native proteins at endogenous expression levels or direct interference with cellular processes [1]. For these compartments, direct expression of the antibodies inside of the same cell ("intrabody" approach; for a review, see Marschall et al. 2015) is of limited applicability because many antibodies do not fold correctly if expressed in the cytosol [2]. The delivery of functional antibody protein to the cytosol from the outside could provide a solution [1, 2]. In spite of the numerous reports claiming successful cytosolic delivery and touting new delivery reagents for proteins, the actual application of protein delivery in research has remained comparably sparse, and publications unambiguously demonstrating the delivery of noteworthy quantities are very rare.

The reason for this might be found in a generally very low efficiency of many of the protein delivery methods published due to overestimation of the amount of protein that has actually ended up in the cytosol. Detection methods which can unambiguously discriminate between outer membrane attachment, endosomal (vesicular) enrichment, or true cytoplasmic delivery are sparse and have not been used frequently. Overestimation of cytosolic delivery further results from artifacts or inappropriate controls. For example, Richard and colleagues found methodological artifacts had led to a marked overestimation of cytosolic delivery of proteins by the so-called cell-penetrating peptides [3]. Furthermore, common misinterpretations originate from the property of many delivery vehicles to attach to the cell surface due to their cationic net charge. For this reason, analysis by flow cytometry alone is not sufficient to assess cytosolic delivery because it does not allow distinguishing surface-bound proteins and those entrapped in endosomes from proteins that have been released to the cytosol [4]. Confocal microscopy has been frequently used, but unambiguous discrimination of antibodies evenly distributed in the cytosol from background is difficult. Many papers claiming detection of cytosolic delivery by microscopy show significant amounts of antibodies in vesicular compartments. Some "positive" control proteins delivered with commercially available protein transfection agents are labeled with fluorescent dyes, which can make them sticky and lead to a high basic signal level of unspecific attachment to cells. A more appropriate control is an unlabeled antibody that can bind to a cytosolic protein with a characteristic pattern such as the cytoskeleton, so surface-bound antibodies cannot be mistaken for a protein delivered to the cytosol (Fig. 1a). The efficiency of protein delivery reagents can be evaluated in this way, and if a positive staining pattern seen in this assay, this also demonstrates the delivery of a considerable amount of functional antibodies, sufficient to bind to the target throughout the cytosol.

An approach to compare protein transfection efficiencies is the use of a reporter system based on Fc-Cre as a cargo protein, a fusion of the Cre recombinase with the constant part of a human antibody [4]. This system requires function of the Cre domain of the delivered protein in the cell and allows identifying only cells where the test protein reached the cytosol (Fig. 1b). In contrast to the assay using antibodies to a characteristic cytoskeleton pattern, the reporter cell assay does not allow conclusions on the quantity of delivered protein per individual cell but determines the fraction of all cells which got proteins delivered to their cytosol at all. Previous studies have shown that protein transfection efficiency shows enormous variations, depending on the transfection reagents used [4].

In the following, we provide the protocols for these two strategies to assess antibody transfection efficiencies. Both unequivocally detect whether antibodies indeed have reached the cytosol as a
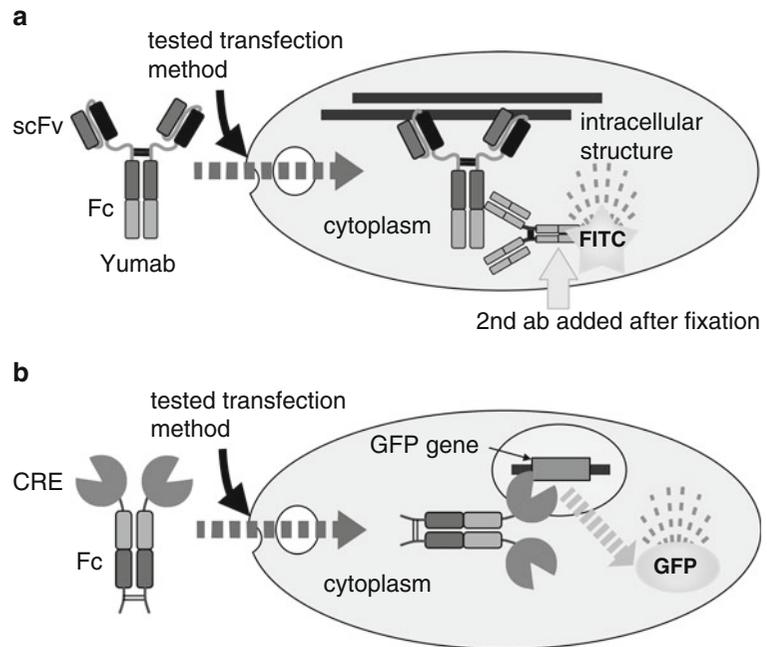
Fig. 1 Two evaluation methods to assess cytoplasmic delivery of protein. (**a**) Staining of typical cytoskeleton proteins. (**b**) Induction of GFP production by *cre* recombinase-dependent DNA recombination. *ab* antibody, *CRE cre* recombinase, *Fc* fragment crystallizable region of a human immunoglobulin G, *FITC* fluorescein isothiocyanate, *GFP* enhanced green fluorescent protein, *scFv* single-chain variable fragment of human immunoglobulin G

result of the application of a delivery reagent/method. Finally, we describe the best antibody delivery method as identified using these two assays in a comparison of various delivery approaches [4].

2 Materials

1. HeLa cells (cell line derived from a human cervix carcinoma ATCC no. CCL-2) (*see* **Note 1**).

2. SC1 REW22 Cre indicator cell line based on the murine fibroblast cell line SC1 [5] (*see* **Note 1**).

3. Cargo proteins. Anti-tubulin (clone F2C) and anti-myosin (clone SF9) scFv-Fc antibodies [6] or Fc-Cre [4] can be produced in HEK293-6E suspension cells according to the protocols by Schirrmann and Büssow [7] and Jäger et al. [8] (*see* **Note 2**).

4. Tissue culture plates (6-, 12-, 24-, 96-wells), Greiner and SPL Life Sciences, Germany.

5. Glass coverslips (15 mm diameter #1, 0.13–0.16 mm thickness) for microscopy from Menzel Gläser (Braunschweig, Germany).

6. Object slides (76 × 26 mm ISO 8037/1) for microscopy from Menzel Gläser (Braunschweig, Germany).

7. FITC-conjugated anti-human IgG (Fc gamma-specific) antibody (Dianova 109-095098).

8. PBS (170 g/L NaCl, 26.8 g/L Na$_2$HPO$_4$·2H$_2$O, 6.9 g/L NaH$_2$PO$_4$·2H$_2$O).

9. Parafilm.

10. Fluoro-Gel (Electron Microscopy Sciences).

11. Electroporation cuvettes (Gene Pulser/MicroPulser Electroporation Cuvettes 0.4 cm gap cat. No. 1652088 from Biorad).

12. Electropulse Gene Pulser Xcell (BioRAD).

3   Methods

3.1 Evaluation of Cytosolic Protein Delivery Efficiency by Microscopy

1. We employed unlabeled anti-tubulin or anti-myosin scFv-Fc antibodies (yumabs) as a cargo protein for the protein delivery method to be tested (Fig. 1a) (also *see* **Note 3**). IgGs can be used instead as well. Directly labeled antibodies are not suitable since they usually stick unspecifically to living cells and substantially decrease the signal to noise levels. HeLa cells are well suited for microscopy-based detection of protein delivery because their large size allows easy and clear detection of cytoskeletal filaments. As a negative control, incubate cells with the same amount of antibody but without applying the delivery method/reagent, as a control for potential artifacts by surface-attached proteins that may have entered the cytosol during cell fixation.

2. After applying the chosen protein delivery method/reagent, leave cells at 37 °C in the incubator to allow them to readhere in 24-well tissue culture plates (approximately 1–2 × 10^5 cells per well of a 24-well plate). In order to avoid background fluorescence due to unspecific attachment of antibodies or antibody aggregates to the coverslip surface, cells seeded in polystyrene 24-well plates can be reseeded on glass cover slides after settling down for 3 h. To do this, remove the medium from 24-well plates, wash with 500 μl PBS, detach cells with 50 μl trypsin/ EDTA per well, add 150 μl of serum containing medium, and distribute each 100 μl of the cell suspension on two glass coverslips (diameter: 15 mm) that have been placed in the wells of 6-well plates (*see* **Note 4**). The surface tension of the medium keeps the 100 μl drop of medium on the glass coverslips. Avoid adding additional medium to the 6-well plates as glass coverslips tend to float on top of the medium. Incubate reseeded cells overnight before analysis to allow cytoskeletal structures to reorganize in their characteristic pattern.

3. Fixation and staining of cells (*see* **Note 5**). Discard medium from glass cover slides, wash 2× with 1 ml PBS. For detection of myosin-binding antibodies, add 500 µl of 4% formaldehyde and incubate for 15 min at RT (PFA dissolved in PBS at 37 °C in the water bath). Discard formaldehyde, wash with 1 ml PBS, and incubate with 0.1% Triton X (500 µl) for 5 min to permeabilize cells. For the detection of tubulin-binding antibodies, discard medium from glass slides, wash twice with 1 ml PBS, and add methanol (that has been cooled to –20 °C) for 4 min to fix cells. After fixation with methanol, no permeabilization step is required. Wash samples with 1 ml PBS and transfer glass coverslips to a small container lined with parafilm that can be covered in a lightproof way. Add a volume of 100 µl of the detection antibody diluted in PBS (1:100 diluted FITC-conjugated goat anti-human IgG (Fc gamma-specific) antibody from Dianova, cat#109-095-098) and incubate at RT for 1 h protected from light. Wash with 200 µl PBS and stain nuclei by applying 200 µl DAPI (100 µg/ml DAPI in 70% ethanol diluted 1:1000 in PBS) for 2–3 min. Wash twice with 200 µl PBS, wash with 200 µl water, discard remaining liquid from glass slides by slightly tapping the edge of the glass to a tissue paper, and put the glass slide into a drop of Fluoro-Gel on an object slide with cells facing toward the drop of gel. Avoid moving or tilting samples and let the Fluoro-Gel dry in a dark, cool place. Analyze samples with appropriate fluorescence microscopy filters.

Samples should be fixed and stained for analysis in a time range between 24 and 96 h after protein delivery, in which antibodies have been found to be stable in the cytosol and are still not yet too much diluted to reach the detection limit [4].

3.2 Evaluation of Protein Transfection Rates by Flow Cytometry Using Reporter Cells

1. Employ Fc-Cre as a cargo protein and apply your chosen protein delivery method (*see* **Note 3**) to transfect SC1 REW22 (Fig. 1b) reporter cells [5]. As a negative control, incubate cells with the same amount of Fc-Cre but without applying the delivery method.

2. Change medium 24 h after delivery. For analysis, detach cells with trypsin/EDTA and wash cells once by adding 3 ml of FC buffer (0.5% BSA, 5 mM EDTA in PBS), sedimenting cells at $500 \times g$ for 5 min at RT and discarding the supernatant. Resuspend cells in 500 µl FC buffer for analysis, and from now on, keep on ice. Cells can then be analyzed by flow cytometry. In order to ensure sufficient time for allowing Fc-Cre to enter the nucleus and perform recombination to indicate successful protein delivery, it is recommended to analyze delivery efficiency at several time points after delivery. The duration of delivery might differ between different delivery methods. For electroporation of

Fc-Cre, protein delivery was well observable in a time range of 24–96 h after delivery and for the lipid-based profection reagent "PULSin" the same time range was suitable [4].

3.3 Delivery of Proteins to the Cytosol by Electroporation

1. Detach HeLa or SC1 REW22 cells with trypsin/EDTA and wash them to remove trypsin/EDTA and all components of the medium by adding at least 10 ml PBS per 2 ml cell suspension and sedimenting cells at $500 \times g$ for 5 min at RT. After sedimenting, discard supernatant and resuspend cells in an appropriate amount of PBS for counting. Count cells and aliquot 4.4×10^5 cells per sample to a tube in order to sediment them at $500 \times g$ for 5 min at RT. After sedimenting, discard the supernatant and resuspend cells in 400 μl of the protein solution containing the protein to be delivered (*see* **Note 3**). As a control for the maximum signal, a Cre expression plasmid (20–40 μg plasmid DNA per sample) can be electroporated into cells.

2. Transfer these 400 μl of cell suspension to a 0.4 cm electroporation cuvette (BioRad) and apply an electric pulse at 650 μF and 300 V using an exponential pulse type with Gene Pulser Xcell (BioRad). Especially for cells in protein solutions, carefully avoid air bubbles in the electroporation cuvette. After application of the electric pulse, instantly apply 400 μl of prewarmed medium and transfer the cells then to a cell culture dish containing an appropriate amount of prewarmed medium.

3. Depending on the experiment and analysis method, reseed cells or directly analyze cells as described above at different time points after electroporation.

4   Notes

1. Maintain cells in DMEM high glucose from PAA (E15-810) supplemented with 8% FCS (PAA, FCS Gold), 1% sodium pyruvate and 1% penicillin/streptomycin (PAA, cat. no. P11-010).

2. For production of each protein, 100 ml of HEK293-6E cell suspension containing between 1.5 and 3×10^6 cells/ml is recommended to be used for transfection with the expression vectors for anti-tubulin (pFUSE-F2C), anti-myosin (pFUSE-SF9), or Fc-Cre (pCSE5.2-hIgG1Fc-miCre). Proteins should be purified by protein A affinity chromatography as described previously [4]. Proteins are in PBS after purification and should directly be frozen in aliquots and only thawed once directly before use.

3. For protein transfection (profection) based on non-covalent association of antibody and the transfection reagent, it is important to test different ratios of cargo protein and transfec-

tion reagent, because each protein has different biochemical and biophysical properties in contrast to DNA, for which a fixed ratio of DNA to transfection reagent can be used. For protein delivery by electroporation, it is advised to perform electroporation with a series of concentrations (a concentration of approximately 0.3 mg/ml scFv-Fc antibody was optimal in our experiments [4]) in order to avoid either negative results due to insufficient protein concentrations or disturbance of imaging by large aggregates due to extremely high protein concentrations.

4. Adherence of proteins to the cell surface or also to the surface of the solid support may result in considerable disturbance of microscopic analysis. Particularly proteins fused to the HIV-TAT$_{47\text{-}57}$ peptide, a so-called cell-penetrating peptide, leads to an even and complete coverage of the glass surface because this peptide is positively charged and can therefore electrostatically interact with the negatively charged glass surface. Particularly directly labeled proteins may also result in unspecific protein attachment to the cell surface and other surfaces. Other proteins may aggregate at high concentrations and in this way disturb imaging, which can be countered be reseeding cells.

5. The structure of cytoskeletal filaments can be lost if fixation of these structures was not optimal. To make sure negative results originate from lack of delivery and are not due to failed fixation of microtubule or actin filaments (to which myosin is associated [9]), it is advisable to test the fixation procedure by staining these filaments before performing the actual delivery experiment.

Acknowledgments

References

1. Marschall AL, Frenzel A, Schirrmann T et al (2011) Targeting antibodies to the cytoplasm. MAbs 3:3–16

2. Marschall AL, Dübel S, Böldicke T (2015) Specific in vivo knockdown of protein function by intrabodies. MAbs 7:1010–1035

3. Richard JP, Melikov K, Vives E et al (2003) Cell-penetrating peptides. A reevaluation of the mechanism of cellular uptake. J Biol Chem 278:585–590

4. Marschall AL, Zhang C, Frenzel A et al (2014) Delivery of antibodies to the cytosol: debunking the myths. MAbs 6:943–956

5. Will E, Klump H, Heffner N et al (2002) Unmodified Cre recombinase crosses the membrane. Nucleic Acids Res 30:e59

6. Moutel S, El Marjou A, Vielemeyer O et al (2009) A multi-Fc-species system for recombinant antibody production. BMC Biotechnol 9:14

7. Schirrmann T, Büssow K (2010) Transient production of scFv-Fc fusion proteins in mammalian cells. In: Stefan Dübel RK (ed) Antibody engineering. Springer, New York, pp 387–398

8. Jäger V, Büssow K, Wagner A et al (2013) High level transient production of recombinant anti-bodies and antibody fusion proteins in HEK293 cells. BMC Biotechnol 13:52

9. Clarke M, Spudich JA (1977) Nonmuscle contractile proteins: the role of actin and myosin in cell motility and shape determination. Annu Rev Biochem 46:797–822

Chapter 15

Validation of Biomarker Proteins Using Reverse Capture Protein Microarrays

Catherine Jozwik, Ofer Eidelman, Joshua Starr, Harvey B. Pollard, and Meera Srivastava

Abstract

Genomics has revolutionized large-scale and high-throughput sequencing and has led to the discovery of thousands of new proteins. Protein chip technology is emerging as a miniaturized and highly parallel platform that is suited to rapid, simultaneous screening of large numbers of proteins and the analysis of various protein-binding activities, enzyme substrate relationships, and posttranslational modifications. Specifically, reverse capture protein microarrays provide the most appropriate platform for identifying low-abundance, disease-specific biomarker proteins in a sea of high-abundance proteins from biological fluids such as blood, serum, plasma, saliva, urine, and cerebrospinal fluid as well as tissues and cells obtained by biopsy. Samples from hundreds of patients can be spotted in serial dilutions on many replicate glass slides. Each slide can then be probed with one specific antibody to the biomarker of interest. That antibody's titer can then be determined quantitatively for each patient, allowing for the statistical assessment and validation of the diagnostic or prognostic utility of that particular antigen. As the technology matures and the availability of validated, platform-compatible antibodies increases, the platform will move further into the desirable realm of discovery science for detecting and quantitating low-abundance signaling proteins. In this chapter, we describe methods for the successful application of the reverse capture protein microarray platform for which we have made substantial contributions to the development and application of this method, particularly in the use of body fluids other than serum/plasma.

Key words Proteomics, Reverse capture protein microarray, Antibodies, Body fluids, Biomarkers, Bioinformatics

1 Introduction

Proteomic studies to date have underscored the fact that some information in the proteome cannot be predicted simply from its related nucleic acid sequence. With the accumulation of vast amounts of DNA sequences in databases, researchers are realizing that merely having complete sequences of genomes is not sufficient to elucidate biological function. A number of alterations or modifications can occur at translational and posttranslational levels to affect protein function. There is no strict linear relationship
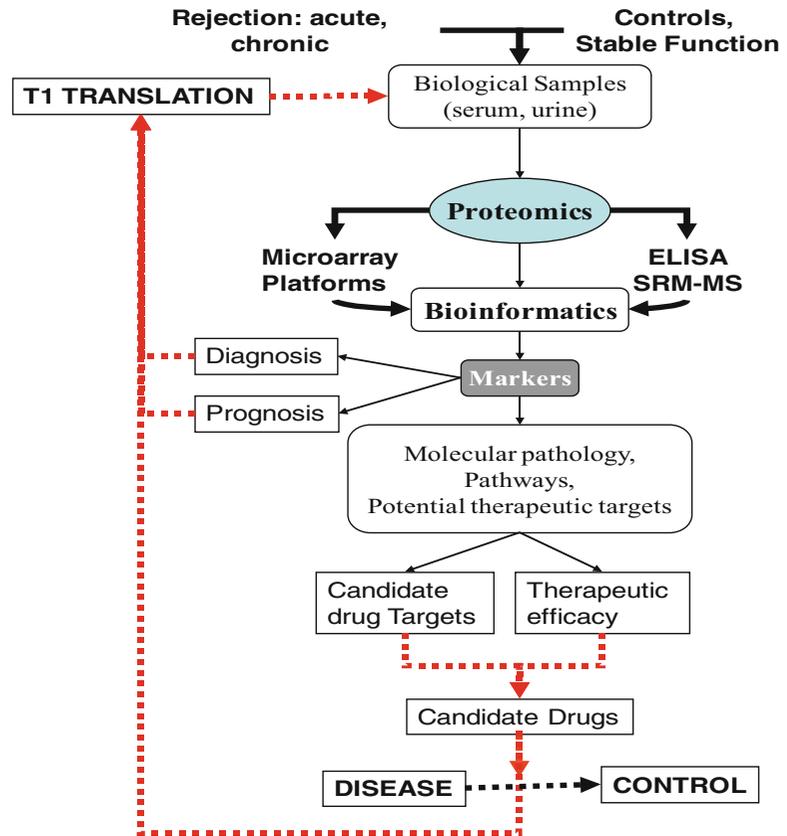
Fig. 1 Workflow for integrative clinical proteomics and translational science for graft rejection

between genes and the protein complement or "proteome" which are the active agents in cells. Importantly, a large number of laboratory tests used in the clinical setting are targeted at expressed proteins. Proteomic analysis also directly contributes to drug development as almost all drugs are directed against proteins. Therefore, there is an urgent need for function-based assays to prioritize and validate candidate targets. Although proteomic studies of any disease are still in their infancy, proteomics is beginning to impact human disease diagnosis and therapeutic intervention.

The promise of clinical proteomics is to make sense of the increasing layers of complexity, by building hypothesis-based insight into a progressive sequence of genomics, proteomics, systems biology, and disease. An example workflow for integrative clinical proteomics and translational science for graft rejection is shown in Fig. 1. Success depends upon the development of quantitative, sensitive, robust, rapid, and massively parallel technologies. It is likely that protein microarrays may make important contributions to this concept and thus personalized medicine. The current tools of proteomic research including two-dimensional

polyacrylamide gel electrophoresis and mass spectrometry have proven to be limited by low throughput and poor sensitivity. Therefore, a strategy that utilizes the existing serum-centric approach in hospitals and clinics for the analysis of biomarker for every phase of the disease is needed to drive clinical decision making, supplementing or replacing currently existing invasive techniques. Protein microarrays are an ideal platform for this approach.

Two primary types of protein array platforms are currently in use. The first is an antibody microarray (forward capture) in which the antibody is printed on the slide and the antigen is present in the solution. The second is reverse capture protein microarrays in which the antigen or protein extract is printed on the slide and the antibody is present in the solution. In this format, a single printed spot on the slide can be viewed as a dot-blot "western." However, the strategy that gives this platform quantitative power is that serial dilutions of body fluids, or cell/tissue extracts can be printed on the slide. Antibody reactivity extinguishes at a given dilution. Extinction permits the calculation of a titer and therefore calculation of a quantitative concentration. The global clinical power of this platform is that hundreds of patient or cell samples can be printed on one slide, thereby allowing a massively parallel analysis to take place. Finally, the global power for this platform is manifest by the ability to print hundreds of identical slides, thereby allowing massively parallel analysis with many antibodies. The reverse capture microarray platform has been used for both cell and tissue extracts [1–4], as well as serum [5, 6]. Using this platform, novel cancer-specific signaling networks were identified for prostate cancer [7, 8], ovarian cancer [9, 10], hematopoietic stem cells [10], NCI-60 cancer cells [3], and primary glioma tissues [11].

An important caveat for these studies is that results have been mostly obtained from limited sets of patients and/or phospho-specific antibodies. However, the results are still encouraging in that they highlight the potential of the microarray-based proteomic approach to distinguish tumor types at the level of signaling networks. The ultimate goal for moving toward patient-tailored therapies is to define such networks and to correlate them with cell-specific responses to specific therapeutic treatments. The proof of concept studies published to date clearly show this potential. Furthermore, as the number of patients, tissue types, and phosphorylated proteins increases, the outcomes may become sufficiently powered to identify universal diagnostic markers and/or to accurately segregate responders versus nonresponders. We have made substantial contributions to the development and application of this method, particularly in the use of body fluids other than serum/plasma [5, 12–14], and in this chapter, we describe methods for the successful application of the reverse capture protein microarray (RCPMA) platform.

2   Materials

2.1 Samples and Sample Preparation

A wide variety of sample types may be used for reverse capture protein microarrays including tissue lysates, serum/plasma, cerebrospinal fluid, urine, bronchoalveolar lavage fluid, etc.

2.1.1 Samples

1. Cell culture pellet or tissue sample (store at −80 °C).
2. Plasma, serum, CSF, or other bodily fluid samples (store at −80 °C).

2.1.2 Sample Preparation

1. T-Per cell lysis buffer (Thermo Fisher Scientific Waltham, MA) (*see* **Note 1**).
2. Protease/phosphatase inhibitors (Halt Inhibitors, Roche Biochemicals, Indianapolis, IN).
3. BCA Protein Assay Kit (Thermo Fisher Scientific Waltham, MA).

2.2 Microarray Printing

1. T-Per cell lysis buffer (Thermo Fisher Scientific Waltham, MA).
2. Extraction/printing buffer (Arrayit Protein Printing Buffer (Arrayit Corporation, Sunnyvale, CA) or 2× SDS sample buffer: 100 mM Tris–HCl, pH 6.8, 4% SDS, 0.2% bromophenol blue, 20% glycerol).
3. 2-Mercaptoethanol or 1 M DTT solution (Sigma Aldrich, St. Louis, MO).
4. Positive control cell lysates or recombinant proteins.
5. Contact array printer (Aushon 2470 Arrayer (Aushon Biosystems, Billerica, MA) or similar printer).
6. Nitrocellulose-coated slides (FAST (Whatman/Sigma Aldrich, St. Louis, MO), ONCYTE (Avid Grace Bio-Labs, Bend, OR) or similar slides).
7. 384-well microwell plates.
8. Desiccant (Drierite, anhydrous calcium sulfate).
9. Ziplock-style plastic storage bags or dessicator.

2.3 Microarray Staining

1. ReBlot Mild Antibody Stripping Solution (EMD Millipore Corporation, Billerica, MA; supplied as a 10× concentrate—dilute to 1× before use).
2. Blocking buffer (5% nonfat dry milk (Santa Cruz Biotechnology, Dallas, TX), 0.1% Tween-20 in 1× Tris-buffered saline (TBS)).
3. Incubation buffer (0.1% BSA, 0.05% Tween20 in 1× TBS).
4. Wash buffer 1 (0.1% Tween-20 in 1× TBS).
5. Wash buffer 2 (1× TBS).
6. Primary antibodies of choice.

7. Fluorescent-labeled secondary antibody.

8. SyproRuby Blot Stain (Life Technologies, Carlsbad, CA).

9. Fixative solution (7 % acetic acid, 10 % methanol).

10. Plastic storage container with tight fitting lid for humidity chamber.

11. Nunc 4-well incubation tray(s) (Thermo Fisher Scientific Waltham, MA).

12. M-Series LifterSlips Cover Slips (Thermo Fisher Scientific Waltham, MA) or other coverslips with raised edge.

2.4 Microarray Analysis

1. Microarray Laser Scanner (Genepix 4000b Laser Scanner (Molecular Devices, Sunnyvale, CA), Innoscan 710-IR (Innopsys, Carbonne, France) or similar scanner).

2. Image analysis software (ImageQuant (GE Healthcare, Pittsburgh, PA) or similar software).

3. Excel (Microsoft, Redmond, WA) or similar spreadsheet software.

3 Methods

3.1 Lysate Preparation

The types of samples that can be used for RCPMA's include tissue samples or cultured cell pellets as well as any proteinaceous body fluid such as serum, plasma, bronchoalveolar lavage fluid, CSF, urine, etc. Tissue samples and cultured cell pellets should be washed with PBS to remove contaminating blood or culture medium. These samples are then lysed in lysis buffer and the protein concentration determined. Most bodily fluids used for RCPMA's should be mixed directly with extraction buffer at the time of printing; these samples are normalized by volume, not total protein concentration.

3.2 Preparation of Tissue or Cell Culture Lysates

1. Add Pierce HALT Protease Inhibitor (1/100 dilution) to lysis buffer immediately before use. Perform all subsequent steps at 4 °C.

2. Add lysis buffer with inhibitors to tissue or cell pellets. Typical amounts are ~1 mL per 200 mg of tissue and ~100 μL per 1×10^6 cultured cells (*see* **Note 2**).

3. Mix cells well with buffer.

4. Disrupt tissue or cell pellets by sonication using short bursts and keeping samples on ice.

5. Centrifuge at $12,000 \times g$, 4 °C for 15 min. Transfer supernatant to new tube.

6. Remove an aliquot for protein assay.

7. Divide lysates into aliquots and freeze at –80 °C.

8. Determine lysate protein concentration using the Pierce BCA Protein Assay Kit (*see* **Note 3**).

3.3 Microarray Printing

RCPMA's allow the determination of levels of specific proteins from hundreds of samples on a single slide, and each batch of printed slides will be interrogated using many antibodies, each with a different affinity constant. To accommodate the range of target protein concentrations and the affinity of each antibody, each sample is printed in a dilution series, permitting selection of those dilution points with signal on the linear portion of the curve for analysis/comparison. For our arrays, we print a series of seven-point, twofold dilutions. Prior to printing, a source plate containing the dilutions for each sample is created in a 384-well plate (*see* **Note 4**) (Fig. 2).

Each array should also include positive and negative controls. Positive controls include a known lysate or recombinant protein mixture that will produce a positive signal with the primary antibody of choice. Negative controls include tissue/cell extracts in which the protein of interest is not expressed as well as printing buffer alone.

Printing is performed using a robotic array printer. Multiple companies manufacture microarray printers; consult the manufacturer's instructions for source plate layout and printhead configuration. We typically print using an Aushon 2470 printer equipped with 16, 185 μm pins in a 4×4 configuration.

3.3.1 Source Plate Creation

Protein microarrays are printed from a source plate, typically a 384-well microwell plate that contains the samples and/or dilutions of the samples. Sample/dilution placement in the source plate is dependent upon the desired microarray layout as well as the microarray printer model and printhead configuration. In our source plates, dilutions for each sample are placed in nonconsecutive wells in 384-well source plate for adjacent placement on the printed microarray (Fig. 2).

3.3.2 Tissue or Cell Culture Lysate Source Plate Preparation

1. Thaw lysates on ice. Perform as many of the following steps as possible on ice or at 4 °C.

2. Determine the volume of diluted lysates required for printing the first dilution points and prepare sufficient 2× printing/extraction buffer by adding protease/phosphatase inhibitors and 2-mercaptoethanol or DTT to a final concentration of 5 % or 50 mM, respectively (*see* **Note 5**).

3. Transfer the appropriate amount of 2× printing/extraction buffer to appropriate wells of 96-well microwell plate for the first dilution.

4. Determine the volume of printing buffer required for printing the subsequent dilution points and prepare sufficient 1× extraction buffer using lysis buffer as the diluent and adding protease/phosphatase inhibitors and 2-mercaptoethanol or DTT to a final concentration of 2.5 % or 25 mM, respectively.

Sample 1   Sample 2   Sample 3

A1	A6	A11	A16	A2	A7	A12	A17	A3	A8	A13	A18	A4	A9	A14	A19	A5	A10	A15	A20	I1	I6	I11	I16
A1	A6	A11	A16	A2	A7	A12	A17	A3	A8	A13	A18	A4	A9	A14	A19	A5	A10	A15	A20	I1	I6	I11	I16
E1	E6	E11	E16	E2	E7	E12	E17	E3	E8	E13	E18	E4	E9	E14	E19	E5	E10	E15	E20	M1	M6	M11	M16
E1	E6	E11	E16	E2	E7	E12	E17	E3	E8	E13	E18	E4	E9	E14	E19	E5	E10	E15	E20	M1	M6	M11	M16

Sample 4   Sample 5   Sample 6

Fig. 2 Example slide layout with source plate positions. Dilutions for each sample are placed in nonconsecutive wells in 384-well source plate for adjacent placement on printed microarray. Alphanumeric designations denote individual wells in microwell plates

5. Transfer 1× printing/extraction buffer to appropriate wells of 96-well microwell plate.

6. Prepare fresh lysis buffer with protease/phosphatase inhibitors, a sufficient amount to dilute tissue or cultured cell lysates to a final protein concentration of 2 mg/mL. Dilute lysates to create normalized 2 mg/mL stock solutions.

7. Dilute the lysate stock solutions ½ into the wells of the microwell plate containing the 2× extraction buffer (*see* **Note 6**).

8. Continue the seven-point, twofold dilution series in the wells containing the 1× extraction buffer. The eighth point contains only printing/extraction buffer as a background control.

9. At this point, the dilution plates may be sealed and stored at –80 °C. If stored, thaw plates, mix well, and centrifuge briefly before proceeding to **step 10**.

10. Incubate sealed plates at 70 °C in heat block or water bath for 15 min. Centrifuge briefly to remove condensation from wall of microwell plates.

11. Transfer 30 μL of each dilution point to the appropriate well of 384-well source plate (*see* **Note 7**), ensuring that no bubbles are produced. Briefly centrifuge plate to remove any bubbles that may have formed.

12. Use source plate immediately (recommended) or seal and store at 4 °C overnight.

3.3.3 Serum/Plasma or Other Body Fluid Source Plate Preparations

1. Thaw samples on ice.

2. Centrifuge at 12,000×*g*, 4 °C for 15 min. Transfer supernatant to a new tube.

3. Determine the volume of diluted samples required for printing the first dilution points and dilute fluid samples appropriately into printing/extraction buffer that contains protease/phosphatase inhibitors and 2-mercaptoethanol or DTT to a final concentration of 5 % or 50 mM, respectively. The final concentration of the printing/extraction buffer should be 1× and 2-mercaptoethanol or DTT to a final concentration of 2.5 % or 25 mM, respectively.

Serum/plasma: Recommended first dilution is 1/10.

Urine, CSF or other fluids: Recommended first dilution is ½.

4. Determine the volume of printing buffer required for printing the subsequent dilution points and prepare sufficient 1× printing/extraction buffer that contains protease/phosphatase inhibitors and 2-mercaptoethanol or DTT to a final concentration of 2.5 % or 25 mM, respectively.

5. Transfer 1× printing/extraction buffer to appropriate wells of 96-well microwell plate.

6. Continue the seven-point, twofold dilution series in the wells containing the 1× extraction buffer. The eighth point contains only extraction buffer as a background control. In contrast to tissue or cell culture lysate dilution plates, *do not* heat serum/plasma samples prior to loading the source plate.

7. Transfer 30 μL of each dilution point to the appropriate well of 384-well source plate (*see* **Note 7**), ensuring that no bubbles are produced. Briefly centrifuge plate to remove any bubbles that may have formed.

8. Proceed immediately to printing; storage of these source plates is not recommended.

3.4 Microarray Printing

Many different microarray printers may be used to print the microarrays. Please refer to the manufacturer's instructions. The following covers the general steps that would apply to any arrayer.

1. Set humidity to 80 %. Fill humidifier with distilled water, if necessary.

2. If possible, set the temperature to 25 °C.

3. Ensure that wash containers have sufficient water and that waste containers are empty.

4. Set print parameters: head configuration, spot spacing, wash parameters, depositions per spot (*see* **Note 8**), and offsets (*see* **Note 9**).

5. Place nitrocellulose-coated slides onto arrayer platen (*see* **Note 10**).

6. Place source plates into source plate "hotel" ensuring that the plastic seal has been removed.

7. Start print run.

8. At the end of the print run, place slides in slides boxes and store with dessicant (in ziplock bags or dessicator) at –20 °C (*see* **Note 11**).

3.5 Array Staining

Reverse capture protein microarrays can be interrogated with any validated antibody (*see* **Note 12**). The primary antibody is detected using a fluorescently labeled secondary antibody. The fluorescently labeled slides are visualized using a laser scanner. For arrays using tissue or cell culture extracts, it is also helpful to use a fluorescent

dye to determine the total protein level in each spot to correct for differences in protein concentration due to sample preparation or printing inconsistencies.

3.5.1   Immunostaining

In addition to the relevant signal arising from the binding of the primary antibody to the specific ligand, the fluorescent signal from any spot often includes a contribution from nonspecific binding of the fluorescent secondary antibody to the sample. Thus, whenever an array set is stained, a negative control slide must be included. The negative control consists of a separate array slide that is incubated with *only* the secondary antibody. For this slide, incubation buffer alone is used during the primary antibody incubation. The resulting nonspecific staining is quantitated and subtracted from the test arrays. Separate controls must be run for mouse and rabbit antibodies. If you have a surplus of slides, it is recommended that you include *two* negative control slides for each secondary antibody, in the event that there is a staining problem with one of the slides.

Prior to staining, many arrays are treated with an antigen retrieval solution to increase the signal-to-noise ratio. *Do not* use antigen retrieval methods on serum/plasma samples or on any array that may contain samples with high levels of albumin as this will cause streaking of the most concentrated spots into the other spots due to the weak binding of albumin to the slide:

1. All incubations are performed with gentle agitation on a rocker.

2. Remove array slides from dessicator and allow them to come to room temperature (~10 min).

3. Incubate slides in 5 mL 1× mild ReBlot solution for 15 min at room temperature. Omit the antigen retrieval step when staining arrays containing plasma/serum samples; proceed directly to **step 5** (Blocking) for serum and plasma arrays.

4. Wash slides 2× with 5 mL of wash buffer 2 for 5 min at room temperature.

5. Incubate slides in 5 mL blocking buffer in incubation chamber for 1 h at room temperature.

6. Wash slides 1× with 5 mL of wash buffer 2 for 5 min at room temperature.

7. While washing, make humidity chamber by placing a wet paper towel in the bottom of the plastic dish.

8. Quickly transfer slides (still in the four-chamber incubation tray) to the humidity chamber, and add 90 μL primary antibody diluted appropriately in incubation buffer to slides. Carefully place coverslip on top of each slide.

9. Place lid on humidity chamber and incubate at 4 °C overnight.

10. Carefully remove coverslips and wash slides 3× with 5 mL of wash buffer 1 for 5 min at room temperature.

11. Incubate slides with fluorescent-labeled secondary antibody (frequently 1:10,000 dilution in 5 mL incubation buffer; *see* **Note 12**) for 1 h at room temperature.

12. Wash slides 3× with 5 mL of wash buffer 1 for 5 min at room temperature.

13. Wash slides 3× with 5 mL of wash buffer 2 for 5 min at room temperature.

14. Dry slides by centrifuging slides in 50 mL conical tubes at $1500 \times g$ for 5 min.

15. Scan slides in laser microarray scanner.

16. Store stained slides at 4 °C, protected from light (foil-wrapped tubes work well).

3.5.2 Total Protein Staining

Slides printed throughout a print run may have slightly different amounts of protein per spot due to evaporation in the source plate. It is best to determine the total protein levels on several slides throughout the print run. At the very least, stain a slide from the beginning of the print run and one from the end of the run.

1. All incubations are performed with gentle agitation on a rocker.

2. Remove array slides from dessicator and allow them to come to room temperature (~10 min).

3. Incubate slide(s) in dH_2O for 5 min at room temperature.

4. Incubate slide(s) in fixative solution for 15 min at room temperature. Use sufficient stain to completely immerse the slide(s).

5. Wash slides 4× with dH_2O for 5 min at room temperature.

6. Incubate slide(s) in SyproRuby stain for 15 min at room temperature. Use sufficient stain to completely immerse the slide (*see* **Note 13**).

7. Wash slides 6× with dH_2O for 5 min at room temperature.

8. Dry slides by centrifuging slides in 50 mL conical tubes at 1500 rpm for 5 min. Do not touch wet slide(s), even while wearing gloves; use forceps to handle wet, stained slide.

9. Scan slides at an excitation wavelength of 532 nm in laser scanner.

10. Store stained slides in foil-covered tubes at 4 °C.

3.6 Microarray Data Analysis

There are several considerations to be taken when analyzing the results of a scanned RCPMA slide including the specific signal arising from the binding of the primary antibody to the protein of interest, as well as background fluorescence and nonspecific binding of the fluorescent secondary antibody to the sample. A typical example of reverse capture protein microarray analysis of stage-specific prostate cancer serum samples is shown in Fig. 3. Local background is usually determined by the scanner software by averaging the signal
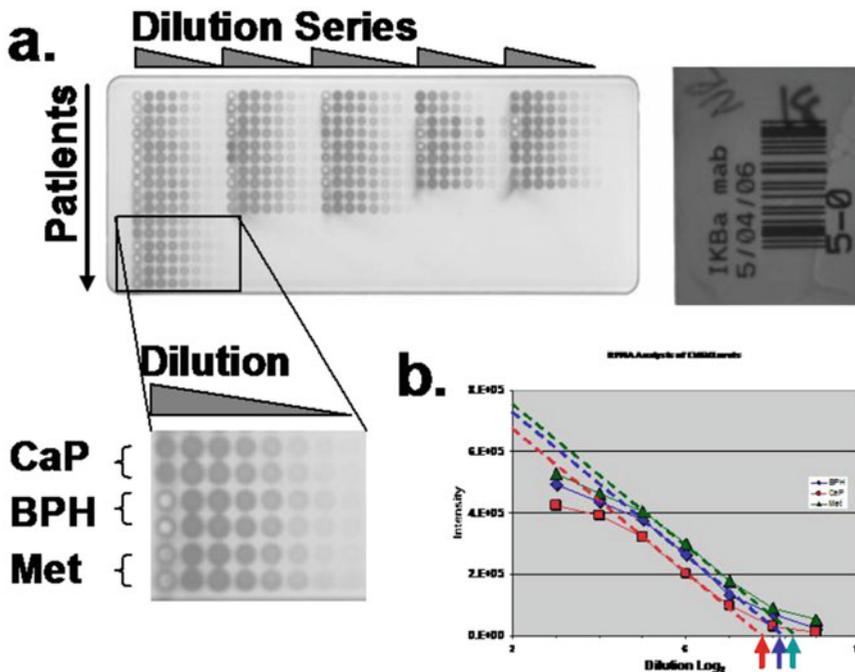
Fig. 3 Reverse capture protein microarray (RCPMA) analysis of stage-specific prostate cancer serum samples. (**a**) RCPMA slide image with an expanded area indicating an eight-point, twofold dilution series of patient samples printed in duplicate. (**b**) Graph of intensity vs. dilution showing the titer determination for patients with benign hyperplasia (*red*), prostate cancer (*blue*), and metastasized prostate cancer (*green*). (Color figure online)

between concentric rings around each spot—the inner ring hugging the edge of the spot, the outer ring drawn a bit beyond the spot edge. Nonspecific binding of the secondary antibody can cause even larger errors in the interpretation of the signal. This signal can be determined from the negative control slide and subtracted from the signal intensity of each spot on the immunostained slide.

In the case of tissue or cell lysate arrays, it is recommended that the measured signals from each spot are corrected for differences in the amount of sample due to sample processing or printing inconsistencies. A slide stained with SyproRuby can be used to correct for the actual amount of total protein printed in each spot. The SyproRuby slide should be analyzed following the same steps detailed for the immunostained slides. The total protein level in each spot can then be used to normalize the signal intensities from the immunostained slide, thereby correcting for inter- and intra-array variability.

Typically, the signal is not linearly proportional to the amount of ligand in the sample throughout the dilution curve. This non-linearity can be corrected by using software such as SuperCurve (http://bioinformatics.mdanderson.org/Software/supercurve/) or circumvented by selecting the linear range or even a single dilution point for comparison across all samples:

1. Use image analysis software to determine the signal and background intensities in each spot. Subtract average background intensity from average signal intensity in each spot. Do this for the slides stained with the primary antibody as well as the negative control slide.

2. Subtract the net intensity for each spot in the negative control slide from the net intensity for each spot in the slide stained with the primary antibody.

3. For arrays containing tissue or cell lysates, normalize each spot to the net intensity obtained for that spot from the SyproRuby total protein stain.

4. Use these normalized, corrected intensities to calculate differences between groups using Excel or other graphing/statistical packages.

4   Notes

1. T-Per is a commonly used commercial lysis buffer, but many homemade lysis buffers are compatible with RCPMA, including RIPA buffer.

2. For tissue, lysis is more complete if the tissue is cut into very small pieces or even ground into "sand" before lysis.

3. Many different protein determination methods may be used; confirm compatibility with lysis buffer components before proceeding.

4. Storage of tissue/cell culture lysate dilution plates at –20 °C is acceptable. However, it is recommended that source plates be used immediately for printing. For fluid samples, neither dilution plates nor source plates should be stored prior to printing.

5. Each well of a 384-well microwell plate easily accommodates 30 μL and this amount will easily print 80–100 slides using 185–350 μm pins.

6. At this point, the sample is in 1× printing/extraction buffer. Further twofold dilutions into 1× printing/extraction buffer made with lysis buffer result in the concentration of buffer components remaining constant while only the protein concentrations change in the dilution series.

7. The source plate layout is dictated by the specific printer and printhead configuration. Please consult the printer manual for additional information.

8. Our typical settings for the Aushon 2470 are:

   4 × 4 printhead configuration

   185 μm pins

   1 s distilled water wash

>500 μm spacing

>350 μm spot size (buffer and sample dependent)

>3–5 depositions per spot (three for concentrated samples like tissue lysates and five for dilute samples such as urine or CSF)

>Triplicate spots

>Two superarrays (replicates of entire array) per slide

9. Top offsets need to be determined for each lot of slides. Dyes (Coomassie or food coloring) may be used in test runs to set offsets for each new lot of slides. Pins must be cleaned thoroughly following print runs using dyes (follow manufacturer's instructions).

10. Handle nitrocellulose-coated slides with gloved hands as the oils can prevent protein binding to the slide.

11. Printed arrays stored at –20 °C have a shelf life of >3 years.

12. Antibodies used for RCPMA should be validated by western blot. Antibodies that produce nonspecific bands are not suitable.

The species reactivity of the fluorescently labeled secondary antibody must be matched with the species of the primary antibody. The fluorescent moiety should match the excitation and emission specifications of the scanner (e.g., AlexaFluor 635 for standard scanners and IRDye 680 for infrared scanners). The use of an infrared scanner substantially limits the autofluorescence of the nitrocellulose, resulting in superior signal-to-noise ratios. Many fluorescent dyes are light sensitive; protect the slide(s) from light during and after staining.

13. SyproRuby stain may be reused up to 4× without a loss in sensitivity. Protect stored stain from light. SyproRuby stain contains both an organic solvent and heavy metals. Do not pour stain down the drain. Keep used stain in a separate waste bottle and dispose according to your institution's regulations.

SyproRuby is light sensitive; protect the slide(s) from light during and after staining.

References

1. Paweletz CP, Charboneau L, Bichsel VE et al (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 20:1981–1989

2. Gulmann C, Espina V, Petricoin E III et al (2005) Proteomic analysis of apoptotic pathways reveals prognostic factors in follicular lymphoma. Clin Cancer Res 11:5847–5855

3. Nishizuka S, Charboneau L, Young L et al (2003) Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. Proc Natl Acad Sci U S A 100:14229–14234

4. Sevecka M, MacBeath G (2006) State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling. Nat Methods 3:825–831

5. Srivastava M, Eidelman O, Jozwik C et al (2006) Serum proteomic signature for cystic fibrosis using an antibody microarray platform. Mol Genet Metab 87:303–310

6. Janzi M, Odling J, Pan-Hammarström Q et al (2005) Serum microarrays for large scale screening of protein levels. Mol Cell Proteomics 4:1942–1947

7. Grubb RL, Calvert VS, Wulkuhle JD et al (2003) Signal pathway profiling of prostate cancer using reverse phase protein arrays. Proteomics 3:2142–2146

8. Sheehan KM, Calvert VS, Kay EW et al (2005) Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. Mol Cell Proteomics 4:346–355

9. Wulfkuhle JD, Aquino JA, Calvert VS et al (2003) Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. Proteomics 3:2085–2090

10. Tibes R, Qiu Y, Lu Y et al (2006) Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol Cancer Ther 5:2512–2521

11. Jiang R, Mircean C, Shmulevich I et al (2006) Pathway alterations during glioma progression revealed by reverse phase protein lysate arrays. Proteomics 6:2964–2971

12. Srivastava M, Eidelman O, Zhang J et al (2004) Digitoxin mimics CFTR-gene therapy, and suppresses hypersecretion of proinflammatory interleukin-8 (IL-8) from cystic fibrosis lung epithelial cells. Proc Natl Acad Sci USA 101:7693–7698

13. Srivastava M, Eidelman O, Torosyan Y et al (2011) Elevated expression levels of ANXA11, integrins β3 and α3, and TNFα contribute to a candidate proteomic signature in urine for kidney allograft rejection. Proteomics Clin Appl 5:311–321

14. Srivastava M, Torosyan Y, Eidelman O et al (2015) Reduced PARP1 as a serum biomarker for graft rejection in kidney transplantation. J Proteomics Bioinform 8:2–6

Chapter 16

Chemical Synthesis of Activity-Based Diubiquitin Probes

Guorui Li, Libo Yuan, and Zhihao Zhuang

Abstract

Activity-based diubiquitin probes are highly useful in probing the deubiquitinase (DUB) activity and ubiquitin chain linkage specificity. Here we describe a detailed protocol to synthesize a new class of diubiquitin DUB probes. In this method, two ubiquitin moieties are connected through a linker resembling the native linkage in size and containing a Michael acceptor for trapping the DUB active-site cysteine. Detailed procedures for generating the linker molecule are also described.

Key words Activity-based probe, Diubiquitin probe, Chemical ligation, DUB, Ubiquitin chain, Linkage specificity

1 Introduction

Deubiquitinases (DUBs) are an important group of proteases that are essential for many cellular processes [1]. Their detection and functional studies have drawn enormous attention in recent years. Among the available strategies, activity-based probes (ABPs) provide useful tools for DUB profiling. A typical DUB probe uses an ubiquitin as the binding or recognition moiety and an electrophilic group introduced at its C-terminus as a warhead to trap the DUBs. The early DUB probes are based on monoubiquitin [2–9]. More recently, this strategy was further developed to generate diubiquitin probes for elucidating the linkage specificity of DUBs [10–14].

We reported a strategy to generate diubiquitin probes [12], in which two ubiquitin moieties are connected through a linker that contains a Michael acceptor for trapping the DUB active-site cysteine. The linker molecule was first ligated to a distal ubiquitin which contained a reactive group at its C-terminus (hUb_{1-75}-MESNa). Then the ubiquitin species with the linker was deprotected and ligated to a mutant proximal ubiquitin to afford the desired diubiquitin probes. Here we describe a detailed protocol for this strategy and also the detailed procedures for the synthesis of the linker molecule.

2   Materials

2.1   Small Molecule Synthesis

Chemical reagents are obtained from Sigma-Aldrich, Alfa, and Acros of the highest available grade and used without further purification. All the organic solvents are purchased from Thermo Fisher Scientific and are certified ACS reagent grade. All the glassware is obtained from Chemglass. Nuclear magnetic resonance spectra are recorded on Bruker AV400 NMR spectrometer with a CryoProbe (^{1}H NMR: 400 MHz, ^{13}C NMR: 100 MHz). Chemical shifts are reported in δ (ppm) units using ^{13}C and residual ^{1}H signals from deuterated solvents as references, and coupling constants J are given in Hz. Peak shapes in NMR spectra are indicated with the symbols "d" (doublet), "s" (singlet), "t" (triplet), and "m" (multiplet). Mass spectra are recorded on Shimadzu LCMS 2020. Analytical thin-layer chromatography (TLC) is performed on silica gel 60 GF254 (Merck). Compounds are visualized under UV254 or by I_2 staining. Column chromatography is conducted on silica gel (230–400 mesh).

2.2   Components for the Synthesis of diUb Probes

All the buffers are prepared using ultrapure water. All the solvents used are HPLC grade. All reagents are prepared and stored at room temperature (unless indicated otherwise). Mass spectra are recorded on Shimadzu LCMS 2020 (monoubiquitin species) or Waters QToF (diubiquitin species) MS instrument equipped with an electrospray ionization (ESI) source:

1. Cleavage buffer for h-Ub$_{1-75}$-MESNa (100 mL): 50 mM MES (2-(N-morpholino)ethanesulfonic acid), 100 mM NaCl, 75 mM β-mercaptoethanesulfonic acid sodium salt (MESNa). Weigh 1 g MES and 0.58 g NaCl and transfer to a 100 mL beaker containing about 50 mL water. Stir until all the solids dissolve. Add water to a volume of 90 mL. Adjust pH to 6.5 with 4 N HCl/NaOH during stirring. Make up to 100 mL with water. Then add 1.23 g MESNa to the above solution. Stir until MESNa dissolves. This buffer should be freshly made.

2. HEPES buffer (1 L): 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 100 mM NaCl, pH 6.7. Weigh 11.9 g HEPES and 5.8 g NaCl and transfer to a 1 L beaker containing about 500 mL of water. Stir until all the solids dissolve. Add water to a volume of 900 mL. Adjust pH to 6.7 with 4 N HCl/NaOH during stirring. Make up to 1 L with water. Store at 4 °C.

3. Dissolving buffer (100 mL): 100 mM Na_2HPO_4, 8 M urea, 500 mM NaCl, pH 6.0. Weigh 1.42 g Na_2HPO_4, 2.9 g NaCl and 48 g urea and transfer to a 100 mL beaker containing about 30 mL of water. Stir until all the solids dissolve (*see* **Note 1**). Add water to a volume of 90 mL. Adjust pH to 6.0 with 4 N HCl/NaOH during stirring. Make up to 100 mL with water.

4. Refolding buffer (1 L, 10×, *see* **Note** 2): 200 mM Na$_2$HPO$_4$, 1 M NaCl, pH 6.0.

5. Buffer A for purification of diUb probe (1 L): 100 mM ammonium acetate, 100 mM NaCl, pH 4.5.

6. Buffer B for purification of diUb probe (1 L): 100 mM ammonium acetate, 1 M NaCl, pH 4.5. Before the purification of diUb probe, buffer A and B are filtered.

3 Methods

3.1 Synthesis of the Linker Molecule 3

The route for the synthesis of the linker molecule **3** is shown in Scheme 1.

3.1.1 Step 1: Synthesis of the Precursor 2

1. The starting material **1** is prepared according to the method of Little et al. [15].

2. The reaction apparatus is set up as shown in Fig. 1 (*see* **Note** 3). A 100 mL round-bottom flask equipped with a Teflon-coated magnetic stir bar is filled with compound **1** (100 mg, 0.32 mmol), ethylene glycol (200 mg, 3.2 mmol), *p*-toluenesulfonic acid (6.2 mg, 0.032 mmol) and 50 mL of benzene (*see* **Note** 4).

3. The flask is placed in silicone oil bath. The reaction mixture is heated to reflux (*see* **Note** 5) and kept at reflux overnight. During this period, the by-product water will be removed by Dean-Stark trap (*see* **Note** 6).

4. The reaction mixture is cooled to room temperature. Then 30 mL saturated NaHCO$_3$ is added to quench the reaction. The biphasic solution is transferred to a 250-mL separatory funnel and layers are separated. The aqueous layer is extracted with diethyl ether (2 × 30 mL).
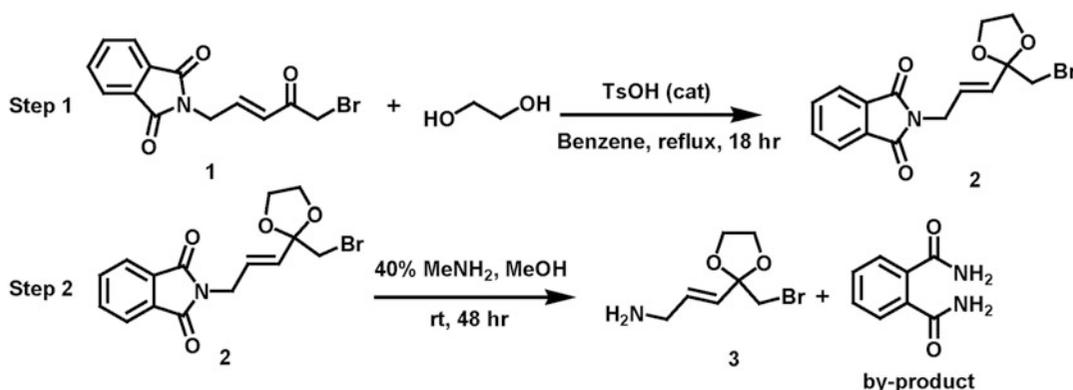
Scheme 1 The route for the synthesis of the linker small molecule **3**. Reagents and conditions: (**a**) *Step 1*: Benzene, *p*-TsOH (cat.), reflux, 18 h, 50 %; (**b**) *Step 2*: MeOH, 40 % MeNH$_2$, rt, 48 h, 77 %
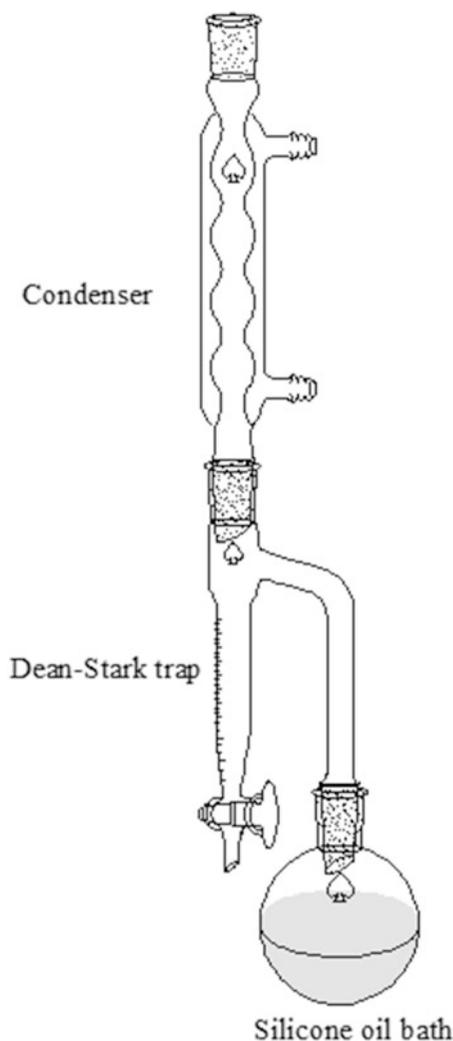
Fig. 1 The reaction apparatus for the synthesis of compound **2**

5. The combined organic layer is washed with H_2O (50 mL), brine (150 mL) and dried with Na_2SO_4. After filtration, the resulting solution is concentrated by rotary evaporation to get the crude mixture as a yellow solid (*see* **Note** 7).

6. The crude mixture is dissolved in dichloromethane (20 mL), mixed with silica gel (1 g) then concentrated by rotary evaporation to get the silica-adsorbed mixture as a yellow powder. Then it is loaded on a column containing 10 g of silica gel and eluted with Hexane/EtOAc (4:1). The fractions containing product are collected and concentrated by rotary evaporation to get compound **2** (57 mg, 0.16 mmol, 50%) as a white solid (*see* **Note** 8).

1. A 100 mL round-bottom flask equipped with a Teflon-coated magnetic stir bar is charged with compound **2** (166 mg, 0.47 mmol), 50 mL of methanol and 2 mL 40 % $MeNH_2$.

2. The neck of the flask is closed by a rubber stopper and the reaction mixture is stirred at r.t. for 48 h under N_2 atmosphere. The solvent is evaporated to get the crude product as brown oil (*see* **Note 9**).

3. The residue is dissolved in dichloromethane (30 mL), mixed with silica gel (1.5 g), and then concentrated by rotary evaporation to get the silica-adsorbed mixture as a yellow powder. Then the powder is loaded on a column containing 15 g of silica gel and eluted with 5 % methanol-NH_3-saturated dichloromethane (*see* **Note 10**). The fractions containing product are collected and concentrated by rotary evaporation to get product **3** (80 mg, 0.36 mmol, 77 %) as yellow oil (*see* **Notes 11** and **12**).

3.2   Synthesis of diUb Probes

The route for the synthesis of diUb probe is shown in Scheme **2**.

3.2.1   Expression and Purification of h-Ub$_{1-75}$-MESNa

The starting material h-Ub_{1-75}-MESNa is expressed and purified according to the reported method [12].

1. The hUb_{1-75}-pTYB1 plasmid is constructed by cloning the humanized yeast Ub gene (lacking Gly76) into the pTYB1 vector (New England Biolabs). The resulting plasmid is confirmed by DNA sequencing.

2. For protein expression, the plasmid is transformed into BL21(DE3) cells. Cells are cultured at 37 °C in LB medium (10 L) containing 100 μg/mL ampicillin. The cell culture is
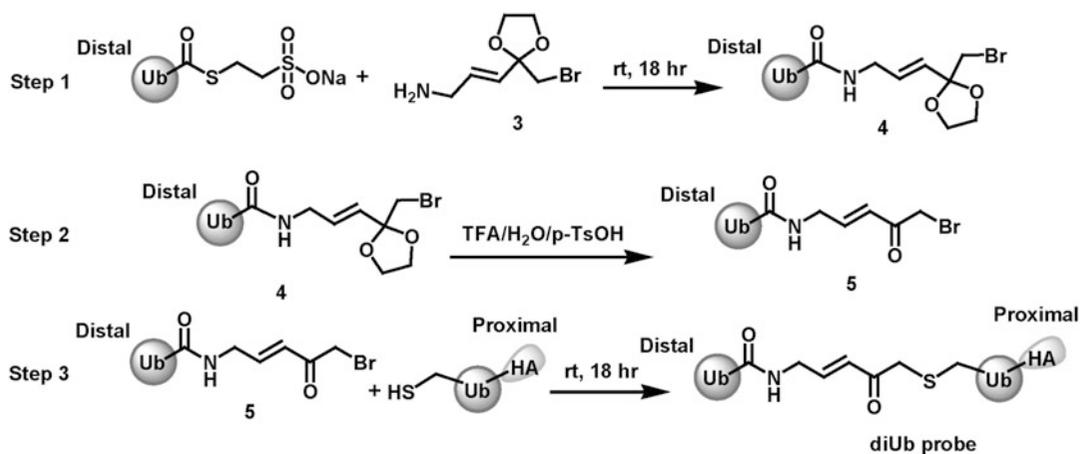
Scheme 2 The route for the synthesis of diUb probes. Reagents and conditions: (**a**) *Step 1*: 3 mg/mL h-Ub_{1-75}-MESNa, 0.4 M Compound **3**, rt, 18 h, 67 %; (**b**) *Step 2*: 0.04 M *p*-TsOH, 54 % (v/v) TFA and 0.5 mg/mL ubiquitin species **4**, rt, 0.5 h, 80 %; (**c**) *Step 3*: rt, 18 h, 63 %

induced with 0.4 mM IPTG at OD600 of 0.6–0.8, and grown for an additional 18 h at 15 °C.

3. Cells are harvested by centrifugation at $8000 \times g$ for 30 min and resuspended in lysis buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 5% glycerol, pH 7.5).

4. Cells are lysed by sonication and the resulting lysate is cleared by centrifugation at $38,000 \times g$ for 30 min.

5. The supernatant is incubated with 50 mL chitin resin (New England Biolabs) at 4 °C for 6 h. The resin is then washed with 500 mL of high-salt wash buffer (20 mM Tris-HCl, 1 M NaCl, 1 mM EDTA, 5% glycerol, pH 7.5) and 300 mL of low-salt wash buffer (50 mM MES, 100 mM NaCl, pH 6.5).

6. The resin is then incubated with 50 mL cleavage buffer (50 mM MES, 100 mM NaCl, 75 mM β-mercaptoethanesulfonic acid sodium salt (MESNa)) for 12 h at room temperature. Then the column is eluted with 50 mL cleavage buffer to get the desired product (*see* **Note 13**).

3.2.2 Step 1: Generation of Ubiquitin Species 4

1. h-Ub$_{1-75}$-MESNa in its original buffer (6 mg/mL, 0.75 mL) is diluted in 10 mL HEPES buffer (50 mM HEPES, 100 mM NaCl, pH 6.7, *see* **Note 14**) and concentrated by a Centricon (MWCO: 3 kDa; volume: 15 mL). The procedure is repeated twice (*see* **Note 15**). Finally it is concentrated to get h-Ub$_{1-75}$-MESNa in HEPES buffer (4 mg/mL, 1 mL).

2. Compound **3** (120 mg) is suspended in 0.33 mL HEPES buffer, and it is vortexed to get a homogeneous mixture. The mixture is added to the solution of h-Ub$_{1-75}$-MESNa to get a solution containing 3 mg/mL h-Ub$_{1-75}$-MESNa and 0.4 M compound **3** (*see* **Notes 16** and **17**). The reaction mixture is rotated at r.t. overnight. MS analysis shows the reaction is complete.

3. The resulting reaction mixture is diluted in 10 mL HEPES buffer and concentrated by Centricon (MWCO: 3 kDa; volume: 15 mL). The procedure is repeated three times (*see* **Note 18**). Finally it is concentrated to get ubiquitin species **4** in HEPES buffer (3 mg/mL, 1 mL, yield 67%).

4. The molecular weight of the product is determined by ESI-MS. Found $M + Na^+ = 8713$ Da, calculated = 8712 Da.

3.2.3 Step 2: Generation of Ubiquitin Species 5

1. The following procedure should be performed on ice. *p*-Toluenesulfonic acid (*p*-TsOH) (42 mg) is dissolved in 1.76 mL H$_2$O, and then it is mixed with 3.24 mL TFA. Ubiquitin species **4** (3 mg/mL, 1 mL) is added to the above solution to get a solution containing 0.04 M *p*-TsOH, 54% (v/v) TFA (trifluoroacetic acid) and 0.5 mg/mL ubiquitin species **4** (*see* **Note 19**).

2. The reaction mixture is rotated at r.t. for 0.5 h (*see* **Note 20**). MS analysis shows the reaction is complete.

3. To the reaction mixture, 30 mL ice-cold diethyl ether is added. A lot of white precipitate forms. Collect the precipitate by centrifugation at (ca. $1600 \times g$) for 5 min. The procedure is repeated twice. The resulting crude mixture is air dried for 10 min.

4. The crude product is dissolved in 3 mL dissolving buffer (100 mM Na_2HPO_4, 8 M urea, 500 mM NaCl, pH 6.0) to get the protein with the concentration of approximately 0.5 mg/mL. Then the solution is transferred into a 3-kDa MWCO dialysis membrane and dialyzed against 1 L refolding buffer (20 mM Na_2HPO_4, 100 mM NaCl, pH 6.0) at r.t. for 6 h (*see* **Note 21**). After dialysis, the product is obtained as a solution in refolding buffer (0.3 mg/mL, 8 mL, yield 80%).

5. The molecular weight of the ubiquitin species **5** is determined by ESI-MS. Found $M + Na^+ = 8669$ Da, calculated $= 8668$ Da.

3.2.4 Step 3: Generation of the diUb Probes

The following protocol is used to generate K48-diUb probe. Other linkages of diUb probe (i.e., K63-diUb probe) can be prepared and purified by the same method as described for K48-diUb probe:

1. In order to form K48-diUb probe, h-K48C-HA-Ub_{1-76} is used as the proximal ubiquitin. h-K48C-HA-Ub_{1-76} is expressed and purified according to the reported method [12].

2. Ubiquitin species **5** (0.3 mg/mL, 8 mL) is mixed with h-K48C-HA-Ub_{1-76} (3 mg/mL, 0.8 mL) and the reaction mixture is incubated at r.t. overnight (*see* **Note 22**).

3. The reaction mixture is analyzed by a 20% denaturing SDS-PAGE gel (Fig. 2).

4. The crude product is purified by a cation-exchange SP Sepharose HP column (GE Life Sciences) using a FPLC system (*see* **Note 23**).

5. The reaction mixture is first buffer exchanged into buffer A (100 mM ammonium acetate, 100 mM NaCl, pH 4.5) using a Centricon (MWCO: 10 kDa; volume: 15 mL).

6. The SP column (5 mL) is pre-equilibrated with buffer A.

7. Then the reaction mixture is loaded onto the SP column and then eluted at a flow rate of 1 mL/min using a gradient of 0–65% buffer B (100 mM ammonium acetate, 1 M NaCl, pH 4.5). Fractions are collected in 1 mL volume. The purification result is analyzed by a 20% denaturing SDS-PAGE gel and fractions containing the pure diUb probe are collected.

8. The pure product is concentrated using a Centricon (MWCO: 10 kDa; volume: 15 mL) to get K48-diUb probe (2.5 mg/mL, 0.6 mL, yield 31%). The SDS-PAGE gel result shows the purity of the product (Fig. 2). And the identity of K48-diUb probe is confirmed by ESI mass spectrometry analysis. Found $M + Na^+ = 18,336$ Da, calculated $= 18,339$ Da.
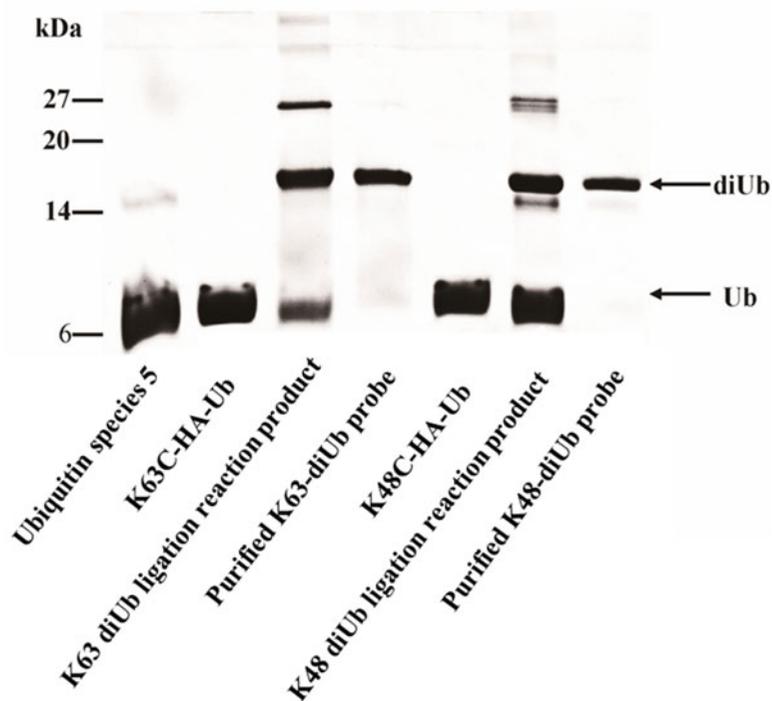
Fig. 2 A 20 % denaturing SDS-PAGE gel result showing the ligation reaction products and the purified diUb probes

4 Notes

1. Urea is difficult to dissolve in water, so it is added in portions and the beaker is heated during stirring. After all the solids dissolve, the beaker is cooled to r.t.

2. Refolding buffer is made and stored as 10× solution. It is diluted to 1× when used.

3. This apparatus is used for a 100 mg scale reaction. For larger-scale reaction, larger apparatus should be used.

4. Care should be taken when using benzene, since benzene is very toxic and a potential carcinogen. All the reaction setup and work-up procedures should be performed in a hood.

5. Benzene can generate an azeotrope with water (boiling point 69 °C) containing 8.8 % water. To ensure the reaction mixture reflux, the oil bath is heated to 100 °C. We also find that toluene can be used instead of benzene, but higher temperature is needed (130 °C).

6. Some cloudy liquid will be produced in Dean-Stark trap, which is the azeotrope of benzene and water.

7. The crude mixture is monitored by TLC analysis, which shows that the starting material is consumed, and a major new compound appears.

8. Analytic data for the pure compound **2** are shown in our previous paper [12].

9. The crude product is monitored by TLC analysis, which is performed with silica gel plates, using 5 % methanol-NH_3-saturated dichloromethane as the eluent. The TLC plates are visualized by both UV254 and I_2 staining. The product is hardly seen under UV254, but it is clearly visible with I_2 staining.

10. NH_3-saturated dichloromethane can be easily obtained by washing dichloromethane with ammonium hydroxide (28–30 % NH_3 in water).

11. Compound **3** is not stable when exposed to the atmosphere at r.t. for a long time. It is suggested that compound **3** is freshly made and then used for next step.

12. Analytic data for the pure product **3** are shown in our previous paper [12].

13. After purification, h-Ub_{1-75}-MESNa is kept in the cleavage buffer containing 50 mM MES, 100 mM NaCl, 75 mM β-mercaptoethanesulfonic acid sodium salt (MESNa). For long time storage, the protein is fast frozen in liquid nitrogen and stored in –80 °C freezer.

14. The pH value of HEPES buffer is very important for this process, because h-Ub_{1-75}-MESNa is very sensitive to pH value. At higher pH (pH > 7), h-Ub_{1-75}-MESNa can be easily hydrolyzed.

15. The aim of this process is to remove MESNa in cleavage buffer, which may affect the ligation of h-Ub_{1-75}-MESNa with compound **3**.

16. The molecular ratio of compound **3**/h-Ub_{1-75}-MESNa is 1150. Excess amount of compound **3** is used to ensure h-Ub_{1-75}-MESNa is converted to ubiquitin species **4**. When too less compound **3** is used, a lot of hydrolysis by-product will be produced.

17. The final concentration of h-Ub_{1-75}-MESNa in the reaction mixture is 3 mg/mL. When its concentration is too high (>4 mg/mL), undesired diUb would be produced.

18. The aim of this process is to remove unreacted compound **3**.

19. Once the protein is added, a lot of white precipitate will appear, but they quickly dissolve after gentle vortex.

20. MS analysis shows 0.5 h is enough for the reaction to be complete. Longer reaction time may produce other by-products.

21. Every 2 h, the folding buffer is changed with fresh one.

22. Our recent result shows that the reaction time can be decreased to 4 h, and it gives similar result as overnight reaction.

23. The purification is based on the fact that diUb binds SP column stronger than monoubiquitin at pH 4.5.

Acknowledgment

References

1. Nijman SM, Luna-Vargas MP, Velds A et al (2005) A genomic and functional inventory of deubiquitinating enzymes. Cell 123:773–786. doi:10.1016/j.cell.2005.11.007

2. Borodovsky A, Kessler BM, Casagrande R et al (2001) A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. EMBO J 20:5187–5196. doi:10.1093/emboj/20.18.5187

3. Borodovsky A, Ovaa H, Kolli N et al (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. Chem Biol 9:1149–1159. doi:10.1016/S1074-5521(02)00248-X

4. de Jong A, Merkx R, Berlin I et al (2012) Ubiquitin-based probes prepared by total synthesis to profile the activity of deubiquitinating enzymes. ChemBioChem 13:2251–2258. doi:10.1002/cbic.201200497

5. Ekkebus R, van Kasteren SI, Kulathu Y et al (2013) On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. J Am Chem Soc 135:2867–2870. doi:10.1021/ja309802n

6. Hemelaar J, Borodovsky A, Kessler BM et al (2004) Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins. Mol Cell Biol 24:84–95. doi:10.1128/MCB.24.1.84-95.2004

7. Love KR, Pandya RK, Spooner E, Ploegh HL (2009) Ubiquitin C-terminal electrophiles are activity-based probes for identification and mechanistic study of ubiquitin conjugating machinery. ACS Chem Biol 4:275–287. doi:10.1021/cb9000348

8. McGouran JF, Kramer HB, Mackeen MM et al (2012) Fluorescence-based active site probes for profiling deubiquitinating enzymes. Org Biomol Chem 10:3379–3383. doi:10.1039/c2ob25258a

9. Ovaa H, Kessler BM, Rolen U et al (2004) Activity-based ubiquitin-specific protease (USP) profiling of virus-infected and malignant human cells. Proc Natl Acad Sci U S A 101:2253–2258. doi:10.1073/pnas.0308411100

10. Haj-Yahya N, Hemantha HP, Meledin R et al (2014) Dehydroalanine-based diubiquitin activity probes. Org Lett 16:540–543. doi:10.1021/ol403416w

11. Iphofer A, Kummer A, Nimtz M et al (2012) Profiling ubiquitin linkage specificities of deubiquitinating enzymes with branched ubiquitin isopeptide probes. ChemBioChem 13:1416–1420. doi:10.1002/cbic.201200261

12. Li G, Liang Q, Gong P et al (2014) Activity-based diubiquitin probes for elucidating the linkage specificity of deubiquitinating enzymes. Chem Commun 50:216–218. doi:10.1039/c3cc47382a

13. McGouran JF, Gaertner SR, Altun M et al (2013) Deubiquitinating enzyme specificity for ubiquitin chain topology profiled by diubiquitin activity probes. Chem Biol 20:1447–1455. doi:10.1016/j.chembiol.2013.10.012

14. Mulder MP, El Oualid F, ter Beek J, Ovaa H (2014) A native chemical ligation handle that enables the synthesis of advanced activity-based probes: diubiquitin as a case study. ChemBioChem 15:946–949. doi:10.1002/cbic.201402012

15. Little TL, Webber SE (1994) A simple and practical synthesis of 2-aminoimidazoles. J Org Chem 59:7299–7305. doi:10.1021/jo00103a021

Chapter 17

Profiling the Dual Enzymatic Activities of the Serine/Threonine Kinase IRE1α

Hannah C. Feldman and Dustin J. Maly

Abstract

There is an allosteric relationship between the kinase and RNase domains of the ER stress sensor IRE1α. This relationship has been exploited to develop ATP-competitive inhibitors that are able to divergently modulate the RNase activity of IRE1α through its kinase domain. Here, we describe a series of biochemical methods for profiling the dual enzymatic activities of IRE1α. These methods can be used to ascertain how ATP-competitive inhibitors affect the kinase activity of IRE1α and for determining whether these ligands allosterically activate or inactivate RNase activity.

Key words Protein kinase, Endoribonuclease, Allostery, IC_{50}, Inhibitor, RNase

1   Introduction

Inositol-requiring enzyme 1 alpha (IRE1α) is an ER sensor protein that is part of the unfolded protein response (UPR) [1, 2]. This multi-domain protein contains an N-terminal lumenal domain, which resides in the ER, that is linked to cytosolic kinase and endoribonuclease domains through a transmembrane linker [3, 4]. Under ER stress, unfolded proteins accumulate in the ER, resulting in the dimerization/oligomerization of IRE1α's lumenal domains. Lumenal domain clustering in the ER facilitates kinase trans-autophosphorylation, which stabilizes an RNase active dimeric state of IRE1α [5, 6]. The active RNase domain of IRE1α cleaves a 26-nucleotide intron from the mRNA encoding the X-box protein 1 (XBP1) transcription factor. Ligation of cleaved XBP1 by the RNA ligase RtcB yields an mRNA that encodes an activated transcription factor that upregulates the expression of proteins that help remediate ER stress [7, 8]. However, under prolonged stress, the substrate specificity of IRE1α's RNase becomes relaxed, causing decay of ER-localized mRNAs, and the upregulation of proapoptotic and pro-inflammatory elements, eventually leading to programmed cell death [9–11]. Due to IRE1α's role in

determining cell fate, there has been interest in the development of small-molecule inhibitors that modulate its function [12].

A number of biochemical studies have shown that a soluble form of IRE1α—referred to as IRE1α*—that consists of the kinase-RNase domain module is able to mimic many of the behaviors of full-length, membrane-bound IRE1α. Like full-length IRE1α, IRE1α* possesses high basal kinase and RNase activity when phosphorylated on its activation loop. Activation loop phosphorylation also stabilizes the dimerization/oligomerization of IRE1α*. A dephosphorylated form of IRE1α*—dP-IRE1*—demonstrates low kinase and RNase activity and is mainly monomeric in solution. The aforementioned properties of IRE1α* and dP-IRE1α* have made these constructs useful tools for studying the biochemistry of IRE1α* and for determining how small molecules are able to modulate the kinase and RNase activities of IRE1α.

There is an allosteric relationship between the kinase and RNase domains of IRE1α, which allows ATP-competitive inhibitors to modulate the RNase activity through the kinase domain [1, 11, 13, 14]. Some ATP-competitive inhibitors are able to activate the RNase domain of IRE1α* and dp-IRE1*, while others—which we call *k*inase *i*nhibiting *R*Nase *a*ttenuators (KIRAs)—are able to inactivate the RNase domain through the kinase domain [1]. In order to systematically probe the effects of ATP-competitive inhibitors on IRE1α, we present a series of experiments utilizing biochemical assays for both the kinase and RNase domains of IRE1α*. Completion of these experiments will allow for the determination of how different ATP-competitive inhibitors affect the kinase and RNase activities of IRE1α, as well as outline methods for generating differentially phosphorylated IRE1α constructs.

2   Materials

All buffers and solutions were prepared using ultrapure water with a resistivity of at least 18 MΩ at 25 °C. Listed components and reagents were purchased from commercial sources unless otherwise noted. Some reagents are light sensitive and should be handled carefully in the dark where noted.

2.1 Assay Components

1. Kinase assay buffer (5×): 250 mM HEPES (pH 7.5), 5 mM MgCl₂, 500 mM NaCl, 0.25 % Tween-20. Make 250 mL and store at room temperature.

2. RNase assay buffer (5×): 250 mM Tris–HCl (pH 7.5), 5 mM MgCl₂, 500 mM NaCl, 0.25 % Tween-20. Make 250 mL and store at room temperature (*see* **Note 1**).

3. Sodium dodecyl sulfate (SDS) (3×): 240 mM Tris–HCl (pH 6.8), 6 % (w/v) SDS, 30 % glycerol, 2.3 M ß-metamercaptoethanol, 0.06 % (w/v) bromophenol blue.

4. Tris-buffered saline Tween (TBST) (1×): 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.1 % (v/v) Tween-20.

5. Dithiothreitol (DTT) (30 mM): Dissolve 4.62 mg of DTT into 1 mL of water and store on ice until use. Make DTT solutions fresh for each experiment.

2.2 Reagents

1. Purified IRE1α* (residues 547–977) expressed from Sf9 insects cells using a Bac-to-Bac Baculovirus Expression System [1].

2. Lambda phosphatase (New England BioLabs).

3. IRE1α* kinase inhibitors in 20 mM DMSO stocks (store at –20 °C).

4. ATP [γP^{32}]: 3000 Ci/mmol; 10 mCi/mL (Perkin Elmer Life Sciences).

5. 5′-Carboxyfluorescein (FAM)- and 3′-Black Hole Quencher (BHQ)-labeled XBP1 single stem-loop mini-substrate (5′FAM-CUGAGUCCGCAGCACUCAG-3′BHQ) (Dharmacon).

6. Bovine serum albumin (BSA) (10 mg/mL): Dissolve 10 mg of BSA into 1 mL of water. Store at –20 °C.

7. β-Mercaptoethanol (BME) (1.43 M): Dissolve 100 μL BME into 900 μL water. Store at –20 °C.

8. Sodium orthovanadate (Na$_3$VO$_4$) (1 M): Dissolve 183.9 mg in 1 mL water. Store at room temperature.

9. Myelin basic protein (MBP) (10 mg/mL): Dissolve 10 mg in 1 mL water. Store at –20 °C.

10. Phosphoric acid (H$_3$PO$_4$) (0.5 %): Dissolve 10 mL H$_3$PO$_4$ into 1990 mL water. Store at room temperature.

11. Ethylenediaminetetraacetic acid (EDTA) (100 mM): Dissolve 292.24 mg of 10 mL water. Store at room temperature.

12. Urea (8 M): Dissolve 4.8 g urea in 10 mL of 1× assay buffer. Store at room temperature.

13. IRE1α primary antibody, IRE1α phospho-Ser724 specific primary antibody, appropriate secondary antibody.

2.3 Special Equipment, Consumables, and Instruments

1. 96-well PCR microplates (Phenix Research).

2. 25 mL pipetting reservoirs (Phenix Research).

3. P81 phosphocellulose ion exchange paper (Reaction Biology).

4. Phosphor Screen (GE Health Sciences).

5. Phosphor Imagining System (GE Typhoon FLA 9000).

6. Black 384-well microplate (Corning).

7. Fluorometer capable of fluorescence excitation at 494 nM and fluorescence emission at 525 nM (Perkin Elmer 2104 Envision Microplate Reader).

3 Methods

3.1 Enzyme Titration of IRE1α's Kinase Activity*

1. Prepare 650 μL of a mixture of the following concentrations: 1.5× kinase assay buffer, 2.87 mM BME, 3 mM Na$_3$VO$_4$, 0.3 mg/mL MBP, 0.075 mg/mL BSA, and 1.5 mM DTT (*see* **Note 2**).

2. Perform a 2× serial dilution of IRE1α* using an initial concentration of ~3 μM IRE1α* in 1.5× kinase assay buffer in wells 1–8 of a 96 well PCR microplate. Well 9 of the plate should contain only 1.5× kinase assay buffer (*see* **Note 3**).

3. Ensure that the total volume of each well is at least 15 μL.

4. Remove ATP [γP^{32}] from freezer and thaw at room temperature (*see* **Note 4**).

5. Calculate the activity of ATP [γP^{32}] using a radiation decay calculator.

6. Prepare 150 μL of 0.04 μCi/mL ATP [γP^{32}] and distribute evenly into wells 1–9 of a PCR microplate.

7. Into a new 96-well PCR plate, pipette 20 μL reaction mixture from **step 1** into wells 1–9 in triplicate (*see* **Notes 5** and **6**).

8. From the microplate from **step 2**, using a multichannel pipette, pipette 5 μL of the enzyme titration into the reaction mixture from **step 7**.

9. Pipette 5 μL ATP [γP^{32}] from **step 6** into the microplate containing your reaction mixture and enzyme.

10. Incubate at room temperature in area outfitted for use with beta-emitting radiation for 3 h (*see* **Note 7**).

11. After 3 h, spot 4.6 μL of each of your wells onto P81 phosphocellulose ion exchange paper.

12. Wash spotted phosphocellulose paper 3× for 10 min with 0.5% phosphoric acid.

13. After washing, air-dry, followed by carefully wrapping in saran wrap (*see* **Note 8**).

14. Place the wrapped phosphocellulose into a phosphor screen and allow exposure for at least 4 h (overnight exposure is preferred).

15. After adequate exposure, imaging using an instrument capable of phosphor imaging (GE Typhoon FLA 9000) (Fig. 1a).

16. Quantitate the blot using imaging software such as ImageQuant.

17. Using graphing analysis software (e.g. GraphPad Prism), determine linear range of IRE1α* (Fig. 1a).
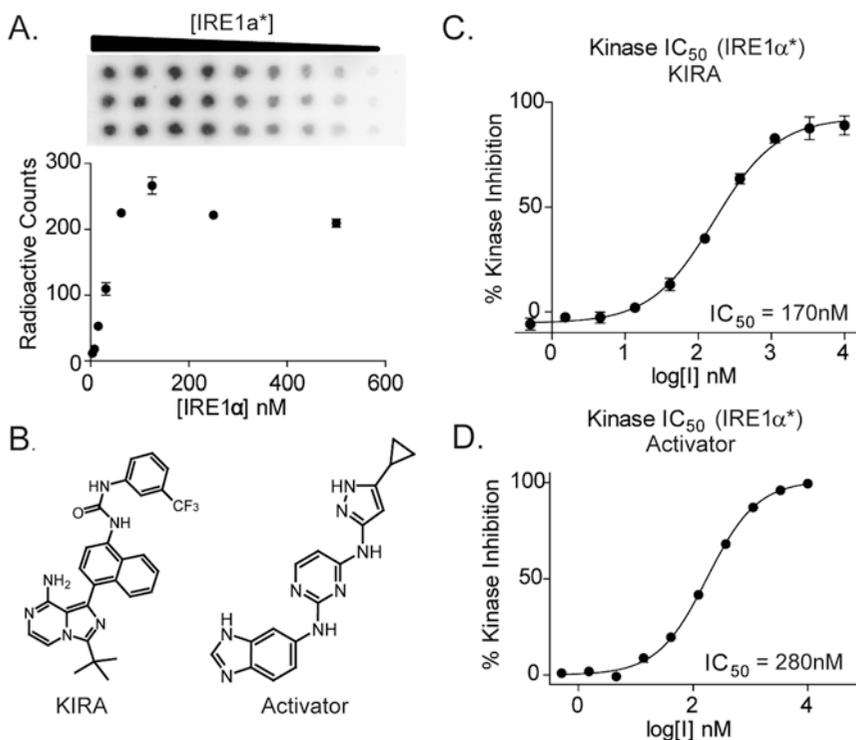
Fig. 1 ATP-competitive inhibitors in the kinase assay. (**a**) *Top*: Titration of IRE1α kinase activity; *bottom*: quantification of the enzyme titration presented above (mean ± S.E.M., *n* = 3). (**b**) Structures of two (KIRA and activator) IRE1α inhibitors characterized in this method. (**c**) IRE1α kinase IC$_{50}$ for the KIRA shown in (**b**) (mean ± S.E.M., *n* = 3). (**d**) IRE1α kinase IC$_{50}$ for the activator shown in (**b**) (mean ± S.E.M., *n* = 3)

3.2 Determining IC$_{50}$ of ATP-Competitive Inhibitors Against IRE1α*'s Kinase Domain

1. Thaw 20 mM inhibitor DMSO stock on ice (Fig. 1b).

2. Perform 3× serial dilutions of inhibitor across wells 1–10. Initial concentration should be in the range of 250–750 μM. Wells 11 and 12 should only contain DMSO (*see* **Note 9**).

3. For each inhibitor titration, prepare 950 μL of a mixture of the following concentrations: 1.26× kinase assay buffer, 0.063 mg/mL BSA, 2.41 mM BME, 2.52 mM Na$_3$VO$_4$, 0.252 mg/mL MBP, 1 mM DTT.

4. Pour into a liquid reservoir of adequate volume.

5. Remove ATP [γP^{32}] from freezer and thaw at room temperature.

6. Calculate the activity of ATP [γP^{32}] using a radiation decay calculator.

7. Prepare 190 μL of 0.04 μCi/mL ATP [γP^{32}] for each inhibitor titration.

8. Distribute evenly into wells 1–12 of a PCR microplate.

9. Pipette 1.2 μL of inhibitor titration in triplicate (wells 1–12, rows 1–3).

10. From reaction mixture reservoir, pipette mixture (without enzyme) into well 12 in each row.

11. In the reaction mixture reservoir, accounting for the subtracted volume from **step 10**, add IRE1α* to the reservoir to a concentration within the linear range determined from the kinase enzyme titration assay above. Mix well.

12. Pipettes 23.8 μL of reaction mixture (with enzyme) into wells 1–11 for each row and allow 30 min incubation time.

13. From the ATP [γP^{32}] from **step 8**, pipette 5 μL into the microplate containing your reaction mixture and enzyme.

14. Incubate, spot, wash, expose, and image as previously described.

15. Determine the percentage of inhibition relative to DMSO treated IRE1α*, and fit with an IC_{50} curve using graphing analysis software (e.g. GraphPad Prism: Non-Linear Regression: One Site IC_{50}) (Fig. 1c and d).

3.3 Enzyme Titration of IRE1α* RNase Activity

1. Prepare a 2× serial dilution of IRE1α* using an initial concentration of ~3 μM IRE1α* in 1.5× RNase assay buffer in wells 1–8 of a 96-well PCR microplate (*see* **Note 10**).

2. Well 9 of the plate should contain only 1.5× RNase assay buffer.

3. Ensure that the total volume of each well is at least 15 μL.

4. Prepare 600 μL 1.5× RNase assay buffer with 2 mM DTT.

5. Prepare 300 μL of a 3 μM solution substrate in 100 μM EDTA.

6. Distribute 20 μL RNase reaction mixture into wells 1–9 of a 384 well black microplate in triplicate (*see* **Note 11**).

7. Add 5 μL IRE1α* titration from **step 1**.

8. Add 10 μL XBP1 substrate from **step 5** and incubate at room temperature in the dark for 10 min (*see* **Note 12**).

9. Quench with 30 μL 8 M Urea to a final concentration of 4 M and a final volume of 60 μL.

10. Spin plate down at $1000 \times g$ for 2 min at room temperature.

11. Image plate using a fluorometer using an excitation wavelength of 494 nm and an emission wavelength of 525 nm.

12. Determine linear range using graphing analysis software (e.g. GraphPad Prism). Determine linear range of IRE1α*.

3.4 Determination of ATP-Competitive Inhibitor Effect Against IRE1α* RNase Domain

1. Prepare a 200 μL of 1.6× RNase assay buffer solution containing 2 mM DTT.

2. Pipette 18.8 μL of buffer from **step 1** in triplicate into a black 384-well microplate.
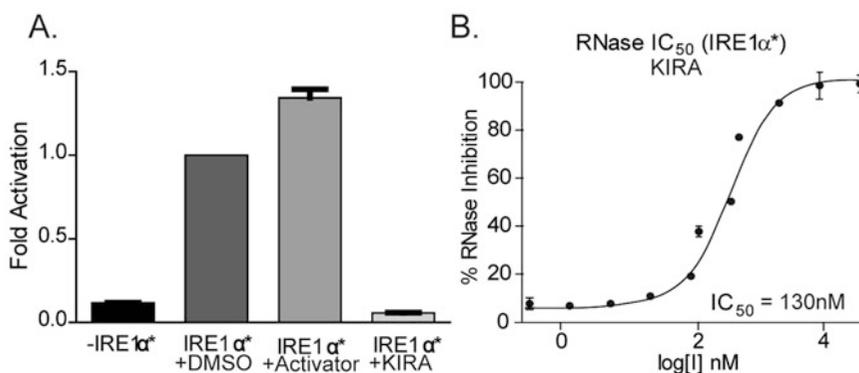
Fig. 2 ATP-competitive inhibitors in the RNase assay. (**a**) Fold activation of IRE1α's RNase activity in the presence of DMSO, 10 μM KIRA, or 10 μM Activator (structures in Fig. 1b) (mean ± S.E.M., $n = 3$). (**b**) RNase IC_{50} for the KIRA shown in Fig. 1b (mean ± S.E.M., $n = 3$)

3. To the remainder of the buffer mix from **step 1**, add IRE1α* to a suitable concentration as determined by the RNase enzyme titration assay above. Pipette 18.8 μL in triplicate into two additional rows.

4. Add 1.2 μL DMSO to the first two rows, and 1.2 μL 250 μM inhibitor of interest into the third row. Incubate for 30 min at room temperature.

5. Prepare 100 μL of a 3 μM solution substrate in 100 μM EDTA.

6. Add 10 μL of substrate into each well and incubate for 10 min.

7. Quench, spin down, and image as described above.

8. Quantitate the fold activation, relative to IRE1α* + DMSO and compare to Fig. 2a.

9. If fold activation resembles an activator from Fig. 2a, dephosphorylate IRE1α* according to procedure 3.5. If fold activation resembles the KIRA from Fig. 2a, determine the IC_{50} of the inhibitor using procedure 3.6 with IRE1α*.

3.5 Dephosphorylation of IRE1α*

1. Incubate IRE1α* with lambda phosphatase (100 units per 0.5 nmol IRE1a*), 50 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM $MnCl_2$, 2 mM DTT, 0.01 % Brij-35 for 1 h at room temperature.

2. Quench reaction by adding Na_3VO_4 to a final volume of 10 mM.

3. Dilute IRE1α* and dP-IRE1α* to a concentration of 500 nM.

4. Dilute in a 2:1 ratio with 3× SDS loading dye. Ensure a final volume of ~10 μL.

5. Run samples in a SDS-PAGE gel (4–15 % Tris–HCl Any kD™ Mini-PROTEAN® TGX Precast Protein Gel; BioRad) and run according to manufacturer's specifications.
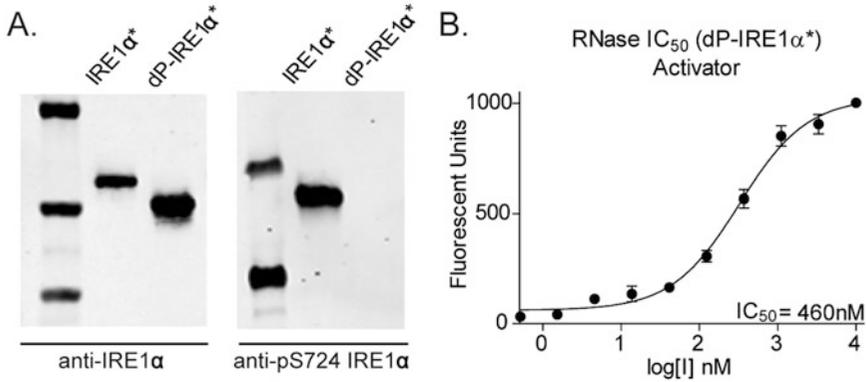
Fig. 3 Dephosphorylation of IRE1α* and Use of dP-IRE1α in RNase Assays. (**a**) *Left*: Western blot of IRE1α* and IRE1α* treated with λ-phosphatase (dp-IRE1α*). Probed with a total-IRE1α antibody; *right*: Western blot of IRE1α* and IRE1α* treated with λ-phosphatase (dp-IRE1α*). Probed with a phospho-Ser724 IRE1α antibody. (**b**) RNase activity of dp-IRE1α* in the presence of the Activator shown in Fig. 1b (mean ± S.E.M., *n* = 3)

6. Visualize extent of dephosphorylation by probing via western blot using total and phosphorylation state-specific IRE1α antibodies (Fig. 3a).

7. If dephosphorylation is successful, make note of the IRE1α* concentration taking into account the dilution from the incubation and quenching. Determine the linear range of IRE1α* before proceeding to inhibitor titration.

3.6 Determination of IC$_{50}$ of ATP-Competitive Inhibitors Against IRE1α* RNase Domain

1. Prepare a 750 µL of 1.6× RNase assay buffer solution containing 2 mM DTT per inhibitor titration.

2. Thaw 20 mM Inhibitor DMSO stock on ice.

3. Preform 3× serial dilutions of inhibitor across wells 1–10. Initial concentration should be in the range of 250–750 µM.

4. Wells 11 and 12 should only contain DMSO.

5. Prepare 400 µL of a 3 µM solution substrate in 100 µM EDTA for each inhibitor titration.

6. Add 1.2 µL inhibitor titration into wells 1–12 in triplicate of a black 384-well microplate.

7. Pipette 18.8 µL reaction mixture into well 12 in each row.

8. Taking into account the subtracted volume from **step 7**, add IRE1α* to a suitable concentration as determined by the RNase Enzyme Titration assay above.

9. Pipette 18.8 µL of IRE1α* containing reaction mixture into wells 1–11 in each row and allow enzyme and inhibitor to incubate at room temperature for 30 min.

10. Add 10 µL XBP1 substrate and incubate at room temperature in the dark for 10 min.

11. Quench, spin down, and image as described previously.

12. Quantitate using data analysis software (e.g., GraphPad Prism) and determine the IC_{50} of the inhibitor (Nonlinear Regression: One Site IC_{50}) (Figs. 2b and 3b) (*see* **Note 13**).

4   Notes

1. RNase assay buffer should be prepared with DNase and RNase free water to prevent non-IRE1α*-mediated mRNA cleavage.

2. Preparing extra reaction mixture is preferred to avoid running out sample in the case of pipetting errors.

3. Enzyme titrations should be performed every time IRE1α* is re-expressed because of varying levels of phosphorylation during expression.

4. ATP $[\gamma P^{32}]$ is radioactive and beta emitting. Proper PPE, safety precautions, should be taken during handled.

5. All assay reactions are run in triplicate in order to report standard error of the mean.

6. Multichannel pipettors are typically used for mixing assay components in order to minimize propagation of error from pipetting.

7. Incubation time may require optimization.

8. Blots can be washed with acetone to expedite the drying process. Imaging wet blots may cause blots to image improperly.

9. Higher initial inhibitor concentration may be preferred for weak inhibitors.

10. Prior to running an assay using mRNA substrate, work area should be clean and wiped down with an RNase decontaminant (Eliminase® Decon Laboratories). All reagents for the assay should be prepared with RNase free water.

11. Any black microplate with low auto-fluorescence may be used.

12. XBP1 mini-substrate is light sensitive and should be handled carefully in the dark.

13. *y*-Axis for activators should be represented by % activation, while *y*-axis for KIRAs will be represented by % inhibition.

References

1. Wang L, Perera BG, Hari SB et al (2012) Divergent allosteric control of the IRE1alpha endoribonuclease using kinase inhibitors. Nat Chem Biol 8:982–989

2. Hetz C, Chevet E, Oakes SA (2015) Proteostasis control by the unfolded protein response. Nat Cell Biol 17:829–838

3. Tirasophon W, Welihinda AA, Kaufman RJ (1998) A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells. Genes Dev 12:1812–1824

4. Wang XZ, Harding HP, Zhang Y et al (1998) Cloning of mammalian Ire1 reveals diversity in the ER stress responses. EMBO J 17: 5708–5717

5. Ali MM, Bagratuni T, Davenport EL et al (2011) Structure of the Ire1 autophosphorylation complex and implications for the unfolded protein response. EMBO J 30:894–905

6. Lee KP, Dey M, Neculai D et al (2008) Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing. Cell 132:89–100

7. Lu Y, Liang FX, Wang X (2014) A synthetic biology approach identifies the mammalian UPR RNA ligase RtcB. Mol Cell 55:758–770

8. Kosmaczewski SG, Edwards TJ, Han SM et al (2014) The RtcB RNA ligase is an essential component of the metazoan unfolded protein response. EMBO Rep 15:1278–1285

9. Han D, Lerner AG, Vande Walle L et al (2009) IRE1alpha kinase activation modes control alternate endoribonuclease outputs to determine divergent cell fates. Cell 138: 562–575

10. Upton JP, Wang L, Han D et al (2012) IRE1alpha cleaves select microRNAs during ER stress to derepress translation of proapoptotic Caspase-2. Science 338:818–822

11. Ghosh R, Wang L, Wang ES et al (2014) Allosteric inhibition of the IRE1alpha RNase preserves cell viability and function during endoplasmic reticulum stress. Cell 158: 534–548

12. Maly DJ, Papa FR (2014) Druggable sensors of the unfolded protein response. Nat Chem Biol 10:892–901

13. Papa FR, Zhang C, Shokat K, Walter P (2003) Bypassing a kinase activity with an ATP-competitive drug. Science 302:1533–1537

14. Feldman HC, Tong M, Wang L et al. (2016) Structural and Functional Analysis of the Allosteric Inhibition of IRE1α with ATP-Competitive Ligands. ACS Chem Biol 11:2195–2205

INDEX