Methods in Molecular Biology 1507

Springer Protocols

Narendra Wajapeyee
Romi Gupta *Editors*

Eukaryotic Transcriptional and Post-transcriptional Gene expression Regulation

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

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

Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation

Edited by

Narendra Wajapeyee and Romi Gupta

Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Humana Press

Editors
Narendra Wajapeyee
Department of Pathology
Yale University School of Medicine
New Haven, CT, USA

Romi Gupta
Department of Pathology
Yale University School of Medicine
New Haven, CT, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-6516-8 ISBN 978-1-4939-6518-2 (eBook)
DOI 10.1007/978-1-4939-6518-2

Library of Congress Control Number: 2016958307

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A

Preface

Regulation of gene expression at the transcriptional and posttranscriptional level is essential for all eukaryotic cells for maintaining their survival and cell identity. Eukaryotic cells have developed intricate and complex control mechanisms that allow them to determine which genes to express and to what extent in a given cell type. Therefore, a comprehensive study of the transcriptional and posttranscriptional mechanisms of gene regulation is expected to provide key insights into almost all of the important biological processes.

This new edition of Methods in Molecular Biology on *Eukaryotic Transcriptional and Posttranscriptional Gene Expression Regulation* compiles a variety of very useful protocols that will allow the reader to study different aspects of transcriptional and posttranscriptional gene expression regulation in eukaryotic cells. These protocols are written in a comprehensive manner to serve as stand-alone protocols, allowing the reader to perform the described method with ease. I anticipate that the researchers in the field of eukaryotic gene expression regulation will find this volume useful.

New Haven, CT

Narendra Wajapeyee
Romi Gupta

Contents

Contributors

Slimane Ait-Si-Ali • *Centre National de la Recherche Scientifique CNRS-Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, Paris, France*

Sophie Beyer • *Centre National de la Recherche Scientifique CNRS-Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, Paris, France*

Sanchita Bhatnagar • *Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA*

Rutger W.W. Brouwer • *Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands*

Sridar V. Chittur • *Department of Biomedical Sciences, School of Public Health, Cancer Research Center, Rensselaer, NY, USA; Center for Functional Genomics, University at Albany-SUNY, Rensselaer, NY, USA*

Lanelle Connolly • *Department of Biochemistry and Biophysics, Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA*

Rebecca S. Devaux • *Department of Biomedical Sciences, School of Public Health, Cancer Research Center, Rensselaer, NY, USA*

Canan Doganli • *Cardiovascular Research Institute, University of California, San Francisco, CA, USA*

Lars Dölken • *Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK; Institut für Virologie und Immunbiologie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany*

Zizheng Dong • *Department of Pharmacology and Toxicology and IU Simon Cancer Center, Indiana University School of Medicine, Indianapolis, IN, USA*

Olivier Elemento • *HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine and Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, USA*

Laura L. Elo • *Turku Centre for Biotechnology, Turku, Finland*

Michael Freitag • *Department of Biochemistry and Biophysics, Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA*

Matteo Forloni • *Department of Pathaology, Yale University School of Medicine, New Haven, CT, USA*

George E. Gentsch • *The Francis Crick Institute, Developmental Biology Laboratory, London, UK*

Eugenia Giannopoulou • *Biological Sciences Department, New York City College of Technology, City University of New York, New York, NY, USA; Arthritis and Tissue Degeneration Program and the David Z. Rosensweig Genomics Research Center, Hospital for Special Surgery, New York, NY, USA*

Michael R. Green • *Howard Hughes Medical Institute and Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA*

Masayuki Hamakawa • *Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Japan*
Daniel Hart • *Cardiovascular Research Institute, University of California, San Francisco, CA, USA*
Erica Hennessy • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Jason Herschkowitz • *Center for Functional Genomics, University at Albany-SUNY, Rensselaer, NY, USA*
Takaaki Hirotsu • *Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Japan; Department of Biology, Graduate School of Sciences, Kyushu University, Fukuoka, Japan*
Thuy Ho • *Department of Pathology, Yale University School of Medicine, New Haven, CT, USA*
Wilfred F.J. van IJcken • *Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands*
Marcus J.O. Johansson • *Department of Molecular Biology, Umeå University, Umeå, Sweden*
Klaus H. Kaestner • *Department of Genetics and Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
Nicola Kearns • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Hyongbum Kim • *Department of Pharmacology and Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, South Korea; Graduate Program of Nano Science and Technology, Yonsei University, Seoul, South Korea; Department of Pharmacology, Yonsei University College of Medicine, Seoul, South Korea*
Marcy Kuentzel • *Center for Functional Genomics, University at Albany-SUNY, Rensselaer, NY, USA*
René Maehr • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Ashish Misra • *Howard Hughes Medical Institute and Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA*
Krishna Mohan Parsi • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Mihwa Na • *Department of Chemistry and Biochemistry, The City College of New York, New York, NY, USA; Biochemistry Ph.D. Program, The City University of New York Graduate Center, New York, NY, USA*
Ljiljana Paša-Tolić • *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
Suresh Ramakrishna • *Graduate School of Biomedical Science and Engineering, Hanyang University, Seoul, South Korea; College of Medicine, Hanyang University, Seoul, South Korea*
Philippe Robin • *Centre National de la Recherche Scientifique CNRS-Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, UMR 7216 CNRS, Paris, France*
Andrzej J. Rutkowski • *Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK*

Kevin Ryan • *Department of Chemistry and Biochemistry, The City College of New York, New York, NY, USA; Biochemistry Ph.D. Program, The City University of New York Graduate Center, New York, NY, USA*
Melissa Sandoval • *Cardiovascular Research Institute, University of California, San Francisco, CA, USA*
Jogender Tushir-Singh • *Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA*
James C. Smith • *The Francis Crick Institute, Developmental Biology Laboratory, London, UK*
Eric Soler • *INSERM U967, CEA/DSV/iRCM, Université Paris-Diderot, Université Paris-Saclay, Fontenay-aux-Roses, France; Laboratory of Excellence GR-Ex, Paris, France; Institute of Molecular Genetics (IGMM), Laboratory of Molecular Hematopoiesis*
Ralph Stadhouders • *Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), Barcelona, Spain*
Tomi Suomi • *Turku Centre for Biotechnology, Turku, Finland*
Lisha Sun • *Department of Pathology, Yale University School of Medicine, New Haven, CT, USA*
David L. Stenoien • *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
Bharathi Suresh • *Department of Pharmacology and Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, South Korea; Graduate Program of Nano Science and Technology, Yonsei University, Seoul, South Korea*
Sean Thomas • *Gladstone Institute of Cardiovascular Disease, San Francisco, CA, USA; Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, USA*
Susana T. Valente • *Department of Immunology and Microbial Sciences, The Scripps Research Institute, Jupiter, FL, USA*
Mirjam C.G.N. van den Hout • *Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands*
Narendra Wajapeyee • *Department of Pathology, Yale University School of Medicine, New Haven, CT, USA*
Si Wu • *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA; Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, USA*
Ji-Ye Yin • *Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, Hunan, China*
Jia Zhang • *Department of Genetics and Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*
Jian-Ting Zhang • *Department of Pharmacology and Toxicology and IU Simon Cancer Center, Indiana University School of Medicine, Indianapolis, IN, USA*
Zhaorui Zhang • *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
Mowei Zhou • *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*

Chapter 1

Fluorescence Reporter-Based Genome-Wide RNA Interference Screening to Identify Alternative Splicing Regulators

Ashish Misra and Michael R. Green

Abstract

Alternative splicing is a regulated process that leads to inclusion or exclusion of particular exons in a pre-mRNA transcript, resulting in multiple protein isoforms being encoded by a single gene. With more than 90 % of human genes known to undergo alternative splicing, it represents a major source for biological diversity inside cells. Although in vitro splicing assays have revealed insights into the mechanisms regulating individual alternative splicing events, our global understanding of alternative splicing regulation is still evolving. In recent years, genome-wide RNA interference (RNAi) screening has transformed biological research by enabling genome-scale loss-of-function screens in cultured cells and model organisms. In addition to resulting in the identification of new cellular pathways and potential drug targets, these screens have also uncovered many previously unknown mechanisms regulating alternative splicing. Here, we describe a method for the identification of alternative splicing regulators using genome-wide RNAi screening, as well as assays for further validation of the identified candidates. With modifications, this method can also be adapted to study the splicing regulation of pre-mRNAs that contain two or more splice isoforms.

Key words Alternative splicing, Genome-wide, RNA interference, Flow cytometry, RBFOX2, Pre-mRNA

1 Introduction

RNA interference (RNAi) has allowed researchers to overcome challenges associated with classical genetic approaches and enabled them to perform high-throughput gene silencing (knockdown) experiments in cells and organisms. Combining the power of genetic screens with phenotypic assays, RNAi screening has made it possible for researchers to identify new genes and/or gene networks involved in regulating critical cellular processes. RNAi is now widely used in high-throughput screens in both basic and applied biology and has allowed researchers to address key questions underlying a wide variety of

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_1,

biological processes, including signal transduction, cell viability, cell or organelle morphology, protein localization and/or function, drug resistance, and alternative splicing [1–6].

A number of genome-wide RNAi libraries have been developed by academic and commercial entities, with newer libraries emerging as our understanding of effective strategies to design and deliver RNAi reagents improves [7]. Readers unfamiliar with RNAi screening strategies are referred to past reviews on assay development and optimization, high-throughput cell-based pooled format RNAi screens [1, 2, 8], arrayed format RNAi screens [1, 9], and in vivo screening [10]. So far, hundreds of large-scale, cell-based, and in vivo RNAi screens have been carried out in *Drosophila melanogaster*, mouse, and human cells. Furthermore, numerous databases are available that support the browsing and analysis of results from these large-scale RNAi screens [11].

The RNAi screen described below is based on a previous publication from our group in which we sought to gain insights into the mechanism of action of the splicing regulator RBFOX2 by performing a genome-wide loss-of-function short hairpin RNA (shRNA) screen to identify factors that, in addition to the RBFOX2 itself, are required for splicing repression [6]. Our screening strategy (Fig. 1) was based upon an experimental system developed by Wang et al. [12] for the identification of exonic splicing silencers (ESSs) from a random sequence pool. This system uses a three-exon mini-gene construct that serves as a reporter for exon silencing (*see* inset to Fig. 1). Exons 1 and 3 of this construct form a complete mRNA encoding green fluorescent protein (GFP), and exon 2 contains a cloning site into which an oligonucleotide can be

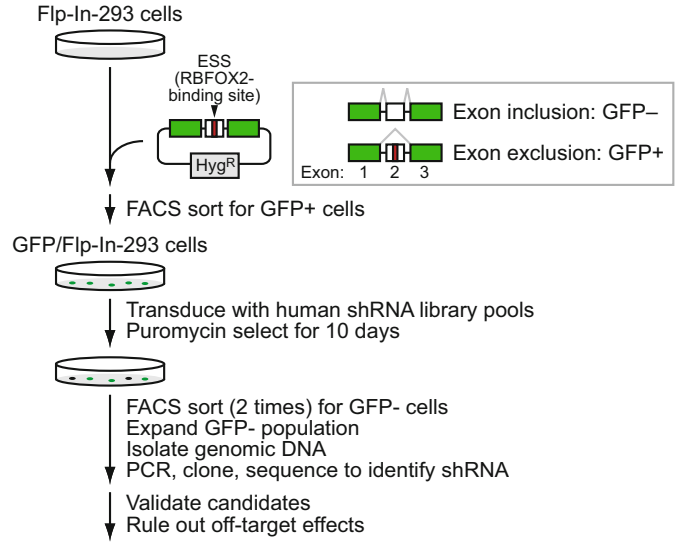

Fig. 1 Schematic of the genome-wide RNAi screening strategy

inserted. Exon 2 is normally included to form an mRNA that does not encode functional GFP. However, insertion of an ESS sequence (in our case, the binding site for RBFOX2) into exon 2 can cause skipping of this exon, producing an mRNA encoding functional GFP. The mini-gene is constructed in an expression vector designed for use with Flp-In™-293 cells, which contain a single Flp recombination target (FRT) integration site. Integration of the mini-gene at a single genomic site is mediated by the Flp recombinase, which is encoded by a plasmid. The cell line containing the stably integrated splicing reporter is first sorted by fluorescence-activated cell sorting (FACS) to obtain a population of cells that is 100% GFP-positive (GFP/Flp-In-293 cells). These cells are then used to perform a genome-wide shRNA screen. Briefly, the GFP/Flip-In-293 cells are stably transduced with an shRNA library; we used The RNAi Consortium (TRC)-Hs1.0 Lentiviral Human Genome shRNA Library comprising ~85,000 shRNAs, which we divided into 22 pools (~5000 shRNAs/pool) to facilitate high-throughput screening. The stably transduced cells from each pool are then FACS sorted to isolate the population of cells in which GFP expression has been significantly diminished and/or lost (GFP-negative), which is the expected result for the loss of splicing repressor function. For each pool, the GFP-negative population of cells is expanded, and the FACS sorting is repeated in order to minimize the number of false positives. The shRNAs in the purified GFP-negative population of cells are identified by sequence analysis. Positive candidates are validated by stably transducing the GFP/Flp-In-293 cells with an individual shRNA directed against the candidate gene and performing FACS analysis as well as other assays using reporter and endogenous target genes.

The method described here is a general screening approach that can be used to identify splicing repressors and/or corepressors. In principle, this screening strategy could also be applied to identify alternative splicing regulators regulating complex alternative splicing events such as the splicing regulation of pre-mRNAs that contain two or more splice isoforms with appropriate modifications to the reporter construct such as the one described by Moore et al. (*see* ref. 5).

2 Materials

Prepare all solutions using ultrapure double-distilled water (ddH_2O). Store all commercially obtained reagents according to the manufacturer's instructions.

2.1 Cell Lines and Culture Conditions

1. Cell lines: Flp-In™-293 cells (Thermo Fisher Scientific) and 293T cells (American Tissue Culture Collection) (*see* **Note 1**).

2. Cell culture medium: DMEM high glucose medium (Invitrogen) containing 10% fetal bovine serum (FBS) (Invitrogen)/Penicillin-Streptomycin (Invitrogen). Mix well and store at 4 °C. Prior to starting the cell culture experiment, warm the media in a 37 °C water bath for about 15 min.

2.2 Lentivirus Preparation, Transduction, and Determination of Multiplicity of Infection

1. 10 cm tissue culture plates.
2. TRC Lentiviral Human Genome shRNA Library (GE Dharmacon) divided into 22 pools, and corresponding positive (RBFOX2) and negative (non-silencing, also called non-targeting) control shRNAs.
3. Lentiviral packaging plasmids pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260).
4. Effectene Transfection Reagent kit (QIAGEN), which includes Effectene reagent, Enhancer, and EC buffer.
5. 0.45 μM filters (Millipore).
6. Polybrene (100 μg/μL) (*see* **Note 2**).
7. Puromycin (5 mg/mL).
8. Phosphate-buffered saline (PBS; 10×): 25.6 g $Na_2HPO_4 \cdot 7H_2O$, 80 g NaCl, 2 g KCl, 2 g KH_2PO_4, ddH_2O to 1 L. Autoclave prior to use. Store at room temperature.
9. Crystal violet staining solution: 40% methanol, 10% acetic acid, 0.01% (w/v) crystal violet in ddH_2O. Store at room temperature.

2.3 Preparation of Stable Cell Lines Carrying the GFP Reporter Construct

1. pcDNA5/FRT vector (Thermo Fisher Scientific) containing the GFP reporter with an RBFOX2-binding site inserted into exon 2 (*see* ref. 12).
2. pOG44 Flp-recombinase expression vector (Thermo Fisher Scientific).
3. Hygromycin B (50 mg/mL) (AG Scientific Incorporation).
4. Cloning cylinders.

2.4 Flow Cytometry Sorting and Analysis

1. Flow cytometer and analyzer, such as a BD FACSCalibur flow cytometer (BD Biosciences).
2. FACS tubes.
3. Collection media: DMEM + 20% FBS/Penicillin-Streptomycin.
4. Trypsin-EDTA (0.25%, Invitrogen).

2.5 Genomic DNA Isolation and Identification of Candidate shRNAs by DNA Sequencing

1. Cell lysis buffer: 0.5% (w/v) SDS, 200 μg/mL of protease K, 10 mM Tris–HCl, pH 8.0, 100 mM NaCl, 10 mM EDTA, pH 8.0. Store at room temperature.
2. Phenol/chloroform/isoamyl alcohol 25:24:1 saturated with 10 mM Tris–HCl, pH 8.0, 1 mM EDTA. Store at 4 °C.

3. Chloroform. Store at room temperature.
4. Sodium acetate (3 M): Dissolve 408.1 g of sodium acetate·$3H_2O$ (MW 136) in 800 mL of ddH_2O. Adjust the pH to 5.2 with glacial acetic acid. Store at room temperature.
5. Ethanol (100% and 70%).
6. TE buffer (1×): 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.
7. Taq PCR buffer (10×): 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris–HCl, pH 8.75, 20 mM $MgSO_4$, 1% Triton X-100, 0.1% BSA. Store at –20 °C. Alternatively, it can be purchased.
8. Taq DNA polymerase.
9. dNTPs (final concentration 10 mM each A, C, G, T).
10. Primers: Primer1 For-TRC (10 μM), TACGATACAAGGC TGTTAGAGAG; Rev-TRC (10 μM), CGAACCGCAAGGA ACCTTC, sequencing primer (MF22; 5 μM), AAACCCAGGGCTGCCTTGGAAAAG.
11. Dimethyl sulfoxide (DMSO).
12. DNase- and RNase-free agarose for gel electrophoresis.
13. QIAquick Gel Extraction Kit (QIAGEN).
14. pGEM®-T Easy Vector Systems kit (Promega), which contains the pGEM®-T Easy Vector, control insert DNA, 2× Rapid Ligation Buffer, and T4 DNA Ligase.
15. DH5α competent cells. Store at –80 °C.
16. 2× LB broth: Dissolve 20 g of peptone, 10 g of yeast extract, and 5 g of NaCl in 1 L of ddH_2O. Autoclave prior to use. Store at room temperature.
17. LB Amp plates: Add 15 g of agar to 1 L of 2× LB broth and autoclave for 25 min. Cool down and add ampicillin (100 μg/mL). Pour into 10 cm dishes, let solidify and store at 4 °C.
18. Isopropyl-ß-D-thiogalactopyranoside (IPTG; 1 M).
19. 5-bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (BCIG or X-gal; 50 mg/mL).

3 Methods

Carry out all cell culture experiments in an ultraviolet-sterilized vacuum hood at room temperature unless otherwise specified. Incubate cells in a 5% CO_2 incubator at 37 °C.

3.1 shRNA Lentivirus Preparation

1. On day 1, plate 2×10^5 293T cells in each of 24 individual 10 cm tissue culture plates; use one plate for each of the 22 shRNA pools, one plate for the positive (RBFOX2) shRNA

control, and one plate for the negative (non-silencing) control shRNA. Shake the plates well to make sure the cells are evenly spread. Incubate at 37 °C for 16 h.

2. On day 2, aspirate old medium and add 10 mL of pre-warmed fresh medium onto the cells. Incubate the cells at 37 °C until the transfection mixture is added. Prepare the transfection mixture by mixing 5 μg of pooled shRNA plasmids (or positive and negative control shRNA plasmids), 2.5 μg of pMD2.G (VSV-G envelope expressing plasmid), and 5 μg of psPAX2 (lentiviral packaging plasmid) in 300 μL of EC buffer. Add 32 μL of Enhancer, mix well by brief vortexing, and let it sit at room temperature for 5 min. Add 80 μL of Effectene, vortex, and let it sit at room temperature for another 20 min. Dispense 0.5 mL of fresh medium to the transfection mixture and, while holding the plate still, gently dispense the entire mixture evenly on top of the cells.
3. On day 3, aspirate all of the medium and add 10 mL of pre-warmed fresh medium. Incubate at 37 °C for 48 h.
4. On day 5, collect the supernatant with a syringe and dispense it through a 0.45 μm filter to remove cell debris. Aliquot the supernatant (1 mL aliquots) into microcentrifuge tubes and store at −80 °C (*see* **Note 3**).

3.2 Determining the Multiplicity of Infection for Lentiviral shRNA Pools

1. Plate 1×10^4 293T cells in each well of a 6-well plate and incubate at 37 °C for ~16 h. Use one plate for each of the 22 pools, plus two more for the positive and negative control shRNAs.
2. Thaw the virus supernatant, and make a series of six tenfold serial dilutions in DMEM media containing 10% FBS/Penicillin-Streptomycin. Mix 100 μL of diluted virus with 900 μL of fresh medium. Add polybrene to a final concentration of 10 μg/mL. Gently dispense the virus mixture on top of the 293T cells and incubate at 37 °C for 24 h.
3. Aspirate the media containing virus and add 10 mL of fresh medium. Incubate at 37 °C for 24 h.
4. Add 1.5 μg/mL of puromycin to each plate and incubate at 37 °C until colonies begin to form (usually about 7–10 days). Change the media containing puromycin every 2 days.
5. Wash colonies with 1× PBS and stain with crystal violet staining solution at room temperature for 20 min. Wash the colonies multiple times with ddH_2O until the water runs colorless. Air-dry the plate and count the colonies. Calculate the multiplicity of infection for the lentiviral supernatants using the following formula:

$$\mathrm{MOI}\left(\text{particle forming units}\left(\text{pfu}\right)/\text{mL}\right) = \text{colony number} \times \text{dilution factor} \times 10.$$

3.3 Preparation of Stable Cell Lines Carrying the GFP Reporter Construct

1. Plate 2×10^6 Flp-In™-293 cells in a 10 cm plate. Incubate the cells at 37 °C for 16 h.
2. Transfect the cells with 2 μg of pcDNA5/FRT-based reporter plasmid and 1 μg of pOG44 plasmid using the Effectene Transfection Reagent kit. Incubate at 37 °C for 24 h.
3. Aspirate all of the medium and add 10 mL of pre-warmed fresh medium.
4. Add hygromycin B (150 μg/mL) (*see* **Note 4**) to enrich cells containing the stably integrated reporter construct. Incubate the cells at 37 °C for ~2 weeks to allow for individual colonies to form. Change the medium containing hygromycin every 4 days. It takes about 8–10 days to wipe out cells that do not carry a stable integration of the construct.
5. Isolate individual colonies (eight to ten) into 6-well plates using cloning cylinders according to manufacturer's instructions.
6. Expand the colonies for 6–8 days in order to obtain enough cells for FACS sorting.
7. Sort the cells using a flow cytometer. Use the parental Flp-In™-293 cells and GFP reporter plasmid-transfected cells as controls to set the gates for the analysis. First, gate for the live cell population in the forward versus side-scatter plot. Next, gate for the GFP-positive cells in the GFP channel: set the gate so that >90% of the cells appear to be GFP-positive in the GFP reporter plasmid-transfected cells and 100% cells appear GFP-negative in the parental Flp-In™-293 cells. Sort all the colonies based on these gates and collect the GFP-positive cells in collection media containing DMEM and 20% FBS/Penicillin-Streptomycin.
8. Plate these cells in 10 cm plates containing DMEM and 10% FBS/Penicillin-Streptomycin. Select the colony that shows maximum mean fluorescence intensity of GFP signal for further experiments.

3.4 shRNA Library Transduction and Selection

1. Plate 2×10^6 GFP/Flp-In-293 cells in 24 individual 10 cm plates, one for each shRNA pool and two for the positive (RBFOX2) and negative (non-silencing) control shRNAs. Incubate at 37 °C for 12–16 h.
2. Transduce the cells with the lentiviral shRNA pools and control shRNAs in a total volume of 10 mL DMEM media containing 10% FBS/Penicillin-Streptomycin and polybrene (10 μg/mL) to achieve a multiplicity of infection (MOI) of 0.2.
3. Change media after 24 h and add puromycin (1.5 μg/mL) to select the cells carrying shRNA. Change media containing puromycin after every 2 days. Usually it takes about 3–4 days to completely wipe out cells that do not contain an integrated shRNA.

3.5 FACS Sorting

1. On day 10 post-infection, aspirate the media from the plates.
2. Rinse the cells with 1× PBS. Add 1 mL 0.25% trypsin to each plate and incubate at room temperature for ~2 min with occasional agitation. Visually inspect the plates to ensure complete detachment of the cells.
3. Add 1 mL of 1× PBS with 10% FBS to neutralize the trypsin and dissociate the cells into a single-cell suspension by repeated pipetting. Collect the cells in FACS tubes and store at 4 °C.
4. Sort the cells using a FACS sorter and analyzer. Use the parental Flp-In™-293 cells and non-silencing shRNA-infected GFP/Flp-In-293 cells as controls to set the gates for FACS sorting. First, gate for the live cell population in the forward versus side-scatter plot. Next, gate for the GFP-positive cells in the GFP channel: set the gate so that >90% of the cells appear to be GFP-positive in the non-silencing shRNA control and 100% cells appear GFP-negative in the parental Flp-In™-293 cells. Sort all the 22 pools based on these gates and collect the GFP-negative cells in collection media containing DMEM and 20% FBS/Penicillin-Streptomycin.
5. Collect the sorted GFP-negative cells from individual pools separately and plate them on a 10 cm dish in media containing DMEM with 10% FBS/Penicillin-Streptomycin and puromycin (1.5 μg/mL). Incubate at 37 °C for 4 days, changing the media containing puromycin every 2 days.
6. On day 17 post-infection, repeat **steps 1–5** and proceed to the next section with the collected GFP-negative cells.

3.6 Genomic DNA Isolation and shRNA Identification

1. Pellet down the cells at 2655 × *g* for 5 min, collect the GFP-negative cells, and resuspend them in 500 μL of cell lysis buffer. Incubate the cell lysate at 55 °C overnight.
2. Add an equal volume of phenol/chloroform/isoamyl alcohol. Mix and centrifuge at 10,621 × *g* for 15 min. Transfer the aqueous phase into a new 1.5 mL microcentrifuge tube and extract again with an equal volume of chloroform.
3. Precipitate the DNA by adding 0.1 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. Mix well by vortexing and leave at −80 °C for at least 1 h. Spin in a tabletop centrifuge at top speed at 4 °C for 30 min, and wash the pellet with 1 mL of 70% ethanol. Pour off the ethanol and invert the microfuge tube onto paper towel to drain the residual ethanol. Air-dry the pellet at room temperature overnight, dissolve it in 100 μL of TE buffer, and measure the DNA concentration (*see* **Note 5**).
4. To amplify the lentiviral shRNA, set up a PCR reaction containing the following components: ~100 ng genomic DNA,

2.5 μL 10× Taq buffer, 1 μL 10 mM dNTPs, 1 μL For-TRC primer, 1 μL Rev-TRC primer, 1 μL DMSO, 0.5 μL Taq DNA polymerase, 18 μL ddH_2O.

5. Program a PCR machine with the following cycling program and run the samples:

Step 1	94 °C for 2 min
Step 2	94 °C for 30 s
Step 3	55 °C for 45 s
Step 4	72 °C for 1 min
Step 5	Go to Step 2 for 34 additional cycles
Step 6	72 °C for 5 min
Step 7	4 °C indefinitely

6. Run the PCR product on a 1% agarose gel containing 10 μl ethidium bromide (10 mg/mL stock). A ~700 bp PCR product should be observed. Elute the product from the gel using a QIAquick Gel Extraction Kit.
7. Ligate the eluted PCR product into the TA cloning vector (pGEM®-T) by setting up a ligation reaction as follows: 3 μL of PCR product, 1 μL of vector, 5 μL of 2× Rapid Ligation buffer, and 1 μL of T4 DNA ligase. Incubate the ligation reaction at 16 °C overnight.
8. The next day, transform the ligation reaction into DH5α competent cells. Plate the transformation mix onto LB Amp plates onto which 10 μL of IPTG and 50 μL of X-gal have been spread evenly.
9. Incubate the plates at 37 °C for ~16 h until the blue and white colonies can be clearly distinguished (*see* **Note 6**).
10. Aliquot 25 μL of ddH_2O into a series of PCR tubes, one for each colony to be picked (*see* **Note 7**). Pick a single white colony from the LB Amp plate using a pipette tip, place the tip in the PCR tube, and mix well by pipetting. Remove 5 μL from each tube, dispense into a fresh PCR tube, and store the remaining 20 μL at 4 °C.
11. Prepare a PCR master mix (by multiplying the following recipe by the number of colonies to be screened) and add 15 μL to each tube prepared for PCR in **step 10**: 2 μL of 10× Taq Buffer, 1 μL dNTPs, 0.5 μL For-TRC primer, 0.5 μL Rev-TRC primer, 0.25 μL Taq DNA polymerase, 10.75 μL ddH_2O.
12. Program a PCR machine with the following cycling program and run the samples:

Step 1	95 °C for 2 min
Step 2	94 °C for 1 min
Step 3	55 °C for 1 min
Step 4	72 °C for 3 min
Step 5	Go to Step 2 for 34 cycles
Step 6	72 °C for 5 min
Step 7	4 °C indefinitely

13. To make sure the PCR reaction worked, load 5 μL of the reaction mixture on a 1 % agarose gel. Again, a ~700 bp PCR product should be observed.
14. Dilute the PCR product by adding 80 μL of ddH_2O to each tube and mix well by pipetting. Mix 2 μL of the diluted PCR product with 2 μL of 5 μM MF22 sequencing primer and send for sequencing.
15. To identify shRNAs from the sequencing results, search for the sequence TTCAAAAA to find the beginning of the shRNA, TCTGAG to define the loop, and CCGGTG to define the end within the sequencing reads. Then map the shRNA sequence onto the TRC shRNA library database (https://www.broadinstitute.org/rnai/trc/lib) to find the corresponding gene.

3.7 Validation of Candidate Genes

1. Prepare individual virus supernatants for each shRNA clone identified from the screen, as described above in Subheading 3.1.
2. Infect 2×10^5 Flp-In™-293 cells with 0.5 mL of virus supernatant mixed with 10 mL of fresh medium and 10 μg/mL of polybrene. Change media after 16 h and then add media containing 1.5 μg/mL of puromycin. Select for 3–4 days.
3. After 10 days, perform FACS analysis in the candidate knockdown cells as described in Subheading 3.5. The analysis should be performed on 1×10^5 cells or more in order to obtain statistically significant results. FACS results obtained for candidate knockdowns should be compared against the control non-silencing shRNA knockdown cells from the same batch. Analyze the flow cytometry data using a software package such as FlowJo software (*see* **Note 8**).
4. Candidates should be validated using other assays, including PCR, to assess changes in isoform abundance of the reporter construct and of known endogenous target genes [6, 12].
5. It is critical to verify that the obtained results are not due to an off-target effect of the shRNA. To do this, select two to three unrelated shRNAs against the same target gene and test whether they confer similar loss of fluorescence signal by FACS and changes in isoform abundance of the reporter gene by

PCR. Candidates that validate with multiple shRNAs can be considered true candidates for follow-up studies. In addition, verify that the candidate shRNAs knock down their target genes with >60–70% knockdown efficiency using quantitative real-time RT-PCR (qRT-PCR) and/or immunoblot analyses (*see* **Note 9**).

4 Notes

1. 293T cells are the preferred cell line for virus preparation due to their high transfection efficiency and their ability to support high expression of virally encoded proteins. The use of early passage 293T cells will ensure high titer virus is obtained.
2. Polybrene improves overall transduction efficiency by enhancing receptor-independent virus absorption through the cell membrane. Pilot experiments should be done to determine the optimum concentration of polybrene required by cells for maximum transduction efficiency. The use of a higher than optimum concentration of polybrene will result in unwanted toxicity to the cells.
3. It is important to note that freeze-thawing will lead to drop in virus titer. Therefore, we recommend avoiding multiple freeze-thaw cycles.
4. It is essential to empirically determine the concentration of hygromycin required for killing the Flp-In™-293 cells. The use of suboptimal concentrations of hygromycin will result in colonies growing on the plate that do not contain stably integrated reporter construct.
5. We use a NanoDrop spectrophotometer (Thermo Scientific) to accurately measure DNA concentrations in solutions of precipitated DNA and PCR reactions.
6. Occasionally, there may be too many colonies growing on the plate due to high competency of DH5α cells. To avoid overcrowding, plate 2 or 3 tenfold serial dilutions of the bacteria on different LB Amp plates containing X-gal and IPTG.
7. The number of colonies to be picked for Sanger sequencing should be estimated based on the percentage of sorted cells collected after FACS. As a starting point, pick 10–20 colonies per pool. Sequencing of colonies from a plate should be stopped when the same shRNA is identified more than 3–4 times.
8. Always use the parental Flp-In™-293 cells and non-silencing shRNA-infected GFP/Flp-In-293 cells as controls when performing FACS sorting and/or analysis with the candidate shRNAs in order to avoid technical variations due to sorting on different days.

9. There are several anticipated classes of "false positives" that one should attempt to eliminate prior to further analysis. For example, shRNAs that result in a decreased level of the splicing repressor or lead to a general reduction of splicing would result in loss of GFP expression and thus score positively in the primary screen. To eliminate such candidates, analyze splicing repressor levels by immunoblotting and qRT-PCR analysis and eliminate those candidates that significantly decrease splicing repressor levels from further experiments.

Acknowledgment

We thank Sara Deibler for editorial assistance. This work was supported by NIH grant R01 GM035490 to M.R.G. M.R.G. is also an investigator of the Howard Hughes Medical Institute.

References

1. Boutros M, Ahringer J (2008) The art and design of genetic screens: RNA interference. Nat Rev Genet 9(7):554–566. doi:10.1038/nrg2364, nrg2364 [pii]
2. Grimm S (2004) The art and design of genetic screens: mammalian culture cells. Nat Rev Genet 5(3):179–189. doi:10.1038/nrg1291, nrg1291 [pii]
3. Bernards R, Brummelkamp TR, Beijersbergen RL (2006) ShRNA libraries and their use in cancer genetics. Nat Methods 3(9):701–706. doi:10.1038/nmeth921, nmeth921 [pii]
4. Mohr S, Bakal C, Perrimon N (2010) Genomic screening with RNAi: results and challenges. Annu Rev Biochem 79:37–64. doi:10.1146/annurev-biochem-060408-092949
5. Moore MJ, Wang Q, Kennedy CJ, Silver PA (2010) An alternative splicing network links cell-cycle control to apoptosis. Cell 142(4):625–636. doi:10.1016/j.cell.2010.07.019, S0092-8674(10)00788-9 [pii]
6. Misra A, Ou J, Zhu LJ, Green MR (2015) Global promotion of alternative internal exon usage by mRNA 3′ end formation factors. Mol Cell 58(5):819–831. doi:10.1016/j.molcel.2015.03.016, S1097-2765(15)00181-1 [pii]
7. Whitehead KA, Langer R, Anderson DG (2009) Knocking down barriers: advances in siRNA delivery. Nat Rev Drug Discov 8(2):129–138. doi:10.1038/nrd2742, nrd2742 [pii]
8. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, Smith Q, Beijersbergen RL, Ghazal P, Shamu CE (2009) Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods 6(8):569–575. doi:10.1038/nmeth.1351, nmeth.1351 [pii]
9. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM (2006) Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat Methods 3(9):715–719. doi:10.1038/nmeth924, nmeth924 [pii]
10. Mohr SE, Perrimon N (2012) RNAi screening: new approaches, understandings, and organisms. Wiley Interdiscip Rev RNA 3(2):145–158. doi:10.1002/wrna.110
11. Mohr SE, Smith JA, Shamu CE, Neumuller RA, Perrimon N (2014) RNAi screening comes of age: improved techniques and complementary approaches. Nat Rev Mol Cell Biol 15(9):591–600. doi:10.1038/nrm3860, nrm3860 [pii]
12. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB (2004) Systematic identification and analysis of exonic splicing silencers. Cell 119(6):831–845. doi:10.1016/j.cell.2004.11.010, S0092867404010566 [pii]

Chapter 2

Tandem Affinity Purification Approach Coupled to Mass Spectrometry to Identify Post-translational Modifications of Histones Associated with Chromatin-Binding Proteins

Sophie Beyer*, Philippe Robin*, and Slimane Ait-Si-Ali

Abstract

Protein purification by tandem affinity purification (TAP)-tag coupled to mass spectrometry analysis is usually used to reveal protein complex composition. Here we describe a TAP-tag purification of chromatin-bound proteins along with associated nucleosomes, which allow exhaustive identification of protein partners. Moreover, this method allows exhaustive identification of the post-translational modifications (PTMs) of the associated histones. Thus, in addition to partner characterization, this approach reveals the associated epigenetic landscape that can shed light on the function and properties of the studied chromatin-bound protein.

Key words Histones, Epigenetics, Chromatin, Post-translational modifications, TAP-tag, Mass spectrometry

1 Introduction

Tandem affinity purification (TAP)-tag approach and subsequent mass spectrometry analysis allow the specific purification of a protein and its interaction partners. It is thus is a useful method to reveal entire protein complexes [1]. Additionally, the purification method described in this chapter allows the separation of cytoplasmic, nuclear soluble, and chromatin-enriched subcellular fractions. The protein of interest, containing a FLAG and HA tags, is either stably overexpressed in a cell line or, alternatively, the endogenous allele(s) can be tagged, thanks to the recent genome editing based [2] to avoid non-specific interactions due to the overexpression conditions.

The TAP-tag approach permits a quantitative and qualitative efficient purification, which is sufficient to detect the interaction partners of the protein of interest. Thus, this purification procedure was extensively used in the past to identify protein-protein interactions and to reveal protein complexes [1].

*Author contributed equally with all other contributors.

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_2,

The preparation of the cellular extracts for TAP-tag consists of two main steps: the separation of cytoplasmic and nuclear fractions and, second, dividing the nuclear fraction into soluble and chromatin-enriched subfractions. This separation additionally enables to identify unexpected interaction partners dependent on the subnuclear compartment [3]. With the focus on chromatin-bound or chromatin-modifying proteins, the complexes recovered from the chromatin-rich fraction could be of main interest. One of main interests here is the identification of the posttranslational modifications (PTMs) of the co-purified histones and of the coprecipitated nucleic acids [4, 5] (Fig. 1).

Thus, the added value of the TAP-tag approach is the possibility to identify PTMs of the purified protein itself and abundant identified interaction partners (originally described in [5]). Thereby, with this feature, the TAP-tag purification is suitable to identify not only new interaction partners but also new enzymatic functions associated with the protein of interest and/or its partners.

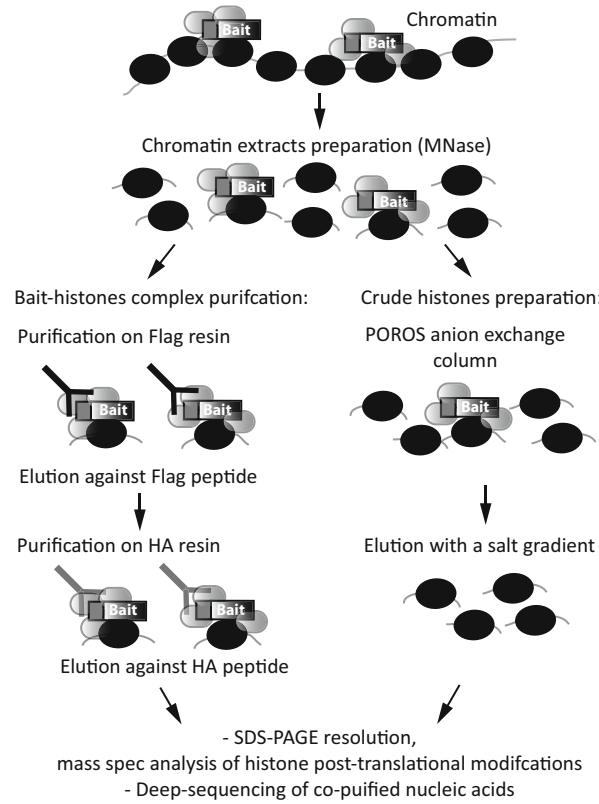

Fig. 1 Post-translational modifications of histones associated with a chromatin-binding protein. Schematic representation of the purification protocol used to purify a bait-histone complexes and crude histones. The latest being used to determine the enrichment compared to the level of a given histone PTM in the input material

In the case of a chromatin-binding protein, this method is thus adapted for identification of the associated "histone code." Indeed, the amino-terminal histone tails, which are exposed on the nucleosome surface, are subject to multiple covalent PTMs. These histone PTMs include lysine and arginine methylation, lysine acetylation, serine and threonine phosphorylation, ADP-ribosylation, ubiquitination, and sumoylation [6]. Histone PTMs confer a unique signature to the nucleosomes involved. Combination of the different modifications on histone N-terminal tails can thus alter chromatin structure to allow gene expression or to repress it, either reversibly or stably. The combinatory pattern of histone PTMs influences the binding and activities of other chromatin-associated proteins that regulate gene expression. Indeed, initial modifications of histones at a specific nucleosome could influence subsequent modifications. Thus, characterizing such modifications associated with a given protein could provide insights into the roles and mechanisms of action of the studied chromatin-binding proteins.

2 Material

2.1 Cells

Cell lines stably expressing FLAG-HA-tagged proteins transduced could be either established using protocol described in [1] or provided by any other mean. Alternatively, the endogenous allele(s) can be tagged using genome editing-based methods [2].

2.2 Buffers

All buffers are used cold if not indicated otherwise and must be supplied with protease inhibitors prior using (*see* **Note 1**).

1. Hypotonic buffer: 10 mM Tris–HCl, pH 7.65, 1.5 mM $MgCl_2$, 10 mM KCl.
2. Low salt buffer: 20 mM Tris–HCl pH 7.65, 25% glycerol, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20 mM NaCl.
3. High salt buffer: 20 mM Tris–HCl pH 7.65, 25% glycerol (12.5 ml), 1.5 mM $MgCl_2$ (75 µl from 1 M), 0.2 mM EDTA, 900 mM NaCl.
4. TEGN: Tris–HCl 20 mM, EDTA 0.1 mM, Glycerol 10%, NaCl 150 mM, NP40 0.01%.
5. Sucrose buffer: 20 mM Tris–HCl, pH 7.65, 15 mM KCl, 60 mM NaCl, 0.34 M sucrose, 0.15 mM spermine, 0.5 mM spermidine.

3 Methods

3.1 Preparation of Cytoplasmic, Nuclear Soluble, and Chromatin-Enriched Fractions

3.1.1 Separating Nuclei and Cytoplasm

1. When working with frozen cell pellets, briefly defrost 20 g cell pellet in a water bath at 37 °C and resuspend first in 10 ml hypotonic buffer. Then add twice 5 ml of fresh hypotonic buffer and hereby wash the pipettes well to obtain the maximum amount of cells. The final volume of the lysate is approximately 40 ml.
2. Use a pre-chilled Dounce homogenizer (40 ml volume) with a tight pestle, and homogenize 20 ml of the lysate with 20 strokes (20 times in and out). Transfer the lysate to a 50 ml tube.
3. Use the second 20 ml lysate and proceed as described in Subheading 3.2.
4. To analyze the efficiency of the lysis, use 30 μl of lysate, mix it with 30 μl of 0.4% trypan blue, and analyze under the microscope. If lysis was efficient, all nuclei are blue. In case of inefficient lysis repeat steps 2 and 3 above.
5. Add 7 ml sucrose buffer (1/3 of the hypotonic buffer volume) supplemented with 0.15 mM spermine and 0.15 mM spermidine. Sucrose buffer preserves the nuclei. Spermine and spermidine avoid leakage by blocking the nuclear pore.
6. Centrifuge the lysate 7 min at 10,000 × *g* to get the nuclei, which are in the pellet. The supernatant is the cytoplasmic fraction.
7. In case of interest in the cytoplasmic fraction (CF), transfer the supernatant as mentioned in above step to a new tube and centrifuge again 7 min at 10,000 × *g*. The supernatant is the CF.

3.2 Preparation of Nuclear Soluble Fraction (NSF) Containing Proteins, Which Weakly Interact with Chromatin

1. Resuspend the nuclei pellet from 1.6 in 10 ml of low salt buffer (one volume equal to the pellet size).
2. Add 10 ml high salt buffer drop by drop while mixing systematically on a vortex. The final concentration of NaCl will be 300 mM.
3. Incubate for 30 min on ice and mix every 5 min.
4. Add 10 ml (1 nuclei pellet volume) of the sucrose buffer.
5. Centrifuge 10 min at 13,000 × *g*. The supernatant is the NSF (*see* **Note 2**).

3.3 Preparation of Nuclear Chromatin-Enriched Fraction (NCF), Containing Proteins Which Strongly Associate with Nucleosomes

1. Resuspend thoroughly the nuclear pellet from 2.5 in 7 ml of sucrose buffer (1 nuclei pellet volume).
2. Add $CaCl_2$ to a final concentration of 1 mM and mix. Starting from a 0.5 M $CaCl_2$ solution, take 28 μl for 14 ml of suspension.
3. Preheat the suspension for 1 min at 37 °C.

4. Add micrococcal nuclease (MNase) to get final concentration of 0.0025 U/µl and mix. Starting from a from 0.5 U/µl stock solution, take 70 µl.
5. Incubate precisely 12 min at 37 °C and mix every 4 min.
6. Immediately place the reaction on ice to stop MNase activity.
7. Add EDTA pH 8.0 to reach 4 mM as final concentration. Starting from a 0.5 M stock solution, take 112 µl EDTA.
8. Perform five cycles of sonication on high amplitude with 1 min each cycle. Between each cycle do 1 min break. The total sonication time will be 10 min.
9. Ultracentrifuge for 30 min at 85,000 × *g*. The supernatant is the NCF.
10. Take 50 µl aliquot of NCF and 100 µl of NSF, which is used as input later, and freeze in liquid nitrogen.

3.4 Protein Complex Purification

3.4.1 Protein Affinity Pulldown by FLAG-Tag

1. Use 600 µl of FLAG affinity resin from the commercial 50% stock for each experimental point (300 µl of pure FLAG resin). Transfer into 15 ml tube, wash with 13 ml of cold TEGN buffer (invert the tube 5 times), centrifuge for 2 min at 1000 × *g*, and remove supernatant. Repeat washing 5 times.
2. Resuspend total amount of FLAG resin in equal volume TEGN buffer, and distribute 600 µl to each experimental point in a 1.5 ml tube (*see* **Note 3**).
3. Incubate over night at 4 °C.
4. Centrifuge 2 min at 1000 × *g* at 4 °C. Keep the supernatant on ice for efficiency check. The FLAG-tagged proteins are in the pellet, bound to the FLAG resin.
5. Resuspend the FLAG resin in 1 ml TEGN buffer and transfer to a 15 ml tube. Repeat this **step 5** times. Use hereby always the same pipette tip for transfer to ensure efficient transfer of all beads.
6. Wash FLAG resin 7 times in the 15 ml tube by adding 13 ml TEGN buffer and inverting the tube 5 times. Do not resuspend the beads with the pipette to avoid losing material. Centrifuge after each washing step for 2 min at 1000 × *g* at 4 °C.
7. Resuspend FLAG resin of each experimental point in 1 ml TEGN buffer and transfer to 1.5 ml tube but keep the 15 ml tube. Centrifuge the 1.5 ml tube (2 min, 1000 × *g*, 4 °C) and remove supernatant.
8. To ensure complete carryover of all beads, rinse the 15 ml tube with 1 ml TEGN buffer, and transfer to the 1.5 ml tube from the previous step. Centrifuge and remove supernatant.
9. Add 200 µl of 4 mg/ml FLAG peptide solution (pH 7.5–8) to the FLAG resin of each experimental point. Add 200 µl TEGN

buffer. Mix by tipping the tube. To avoid losing FLAG resin, do not touch with the pipette tip. Incubate on a rotating wheel overnight (or at least 4 h) at 4 °C.

10. Spin tubes 2 min at 1000 × *g* at 4 °C. Use a flat-narrow pipette tip to transfer the supernatant (FLAG eluate). Do not carry over beads.
11. Centrifuge the FLAG resin again to recover leftover supernatant. Use flat-narrow pipette tips to avoid carryover of beads.
12. To ensure efficient elution from the FLAG resin, take the FLAG beads from **step 11** and repeat **steps 9** and **10**.
13. Combine supernatants from first and second elution.

3.5 Efficiency Test After FLAG Purification

1. Take 15 µl eluate obtained from previous step, add 5 µl of 4× loading buffer and 2 µl 10× reducing agent, and mix (if loading buffer has a different concentration, adjust volumes of buffer and reducing agent).
2. Boil samples 5 min at 95 °C, quick spin samples, and run on a SDS-PAGE.
3. For silver staining use a commercially available silver staining kit and follow the manufacturer's protocol (*see* **Note 4**).

3.6 Protein Affinity Pulldown by HA-Tag

1. Use 300 µl of HA affinity gel from the commercial 50% stock for each experimental point (150 µl of pure HA resin). Transfer into 15 ml tube, wash with 13 ml of cold TEGN buffer (invert the tube 5 times), centrifuge for 2 min at 1000 × *g*, and remove supernatant. Repeat washing 5 times.
2. Resuspend total amount of HA resin in equal volume TEGN buffer and distribute 300 µl to each experimental point in a 1.5 ml tube. Centrifuge 2 min at 1000 × *g* and 4 °C. Eliminate the maximum of the washing buffer by using the flat-narrow pipette tips.
3. Add the eluates from FLAG-based purification (from 3.4.1.13) to the HA resin.
4. Incubate overnight at 4 °C on a rotating wheel.
5. Centrifuge 2 min at 1000 × *g* at 4 °C. Keep the supernatant on ice for efficiency check. The HA-tagged proteins are in the pellet, bound to the HA resin.
6. Resuspend the HA resin in 0.5 ml TEGN buffer and transfer to a new 1.5 ml tube. Repeat this step once and transfer to the same tube. Use hereby always the same pipette tip for transfer to avoid losing beads.
7. Wash HA resin 8 times in the 1.5 ml tube by adding 1 ml TEGN buffer and inverting the tube 5 times. Do not touch the beads with the pipette to avoid losing material. Centrifuge after each washing step for 2 min at 1000 × *g* at 4 °C.

8. Transfer HA resin of each experimental point to a new 0.5 ml tube with the last washing step. Rinse the pipette tip, used for transfer with TEGN buffer, and collect as many beads as possible into the same 0.5 ml tube. Centrifuge the tube (2 min, 1000 × *g*, 4 °C) and remove as much supernatant as possible.
9. Add 100 μl of 4 mg/ml HA peptide solution to the HA resin of each experimental point. To avoid losing HA resin, do not touch with the pipette tip. Incubate on a rotating wheel overnight (or at least 4 h) at 4 °C.
10. Spin tubes 2 min at 1000 × *g* at 4 °C. Use a flat-narrow pipette tip to transfer the supernatant (HA eluate). Do not carry over beads.
11. Centrifuge the HA resin again to recover leftover supernatant. Use flat-narrow pipette tips to avoid carryover of beads.
12. To ensure efficient elution from the HA resin, take the HA beads from **step 11** and perform a second elution. Repeat **steps 9–11**.
13. Combine supernatants from first and second elution.

3.7 Efficiency Test After HA Purification

1. Repeat steps as described in Subheading 3.5.

3.8 Concentration of Eluates

1. Use centrifugal filter units with 10 kDA cutoff.
2. Concentrate the eluate to 30 μl by using the filter unit according to manufacturer's instructions.
3. Take ¼ of the eluate (7.5 μl), snap freeze in liquid nitrogen, and store at −80 °C. It will be used for western blot analysis to confirm the results obtained by mass spectrometry.
4. Take the remaining ¾ (22.5 μl) and prepare samples for mass spectrometry analysis (as in [5, 7]).

3.9 Nucleosomal Histone Preparation for Mass Spectrometry Analysis (See Note 5)

1. Run the purified protein complex from the chromatin-enriched nuclear fraction on a 4–12% acrylamide gradient SDS-PAGE gel.
2. Stain SDS-PAGE gel with Colloidal blue.
3. Cut gel bands corresponding to each histone and destain overnight in 50% acetonitrile, 50 mM NH_4HCO_3.
4. Subject histones to a propionylation-based modification method to study lysine modifications [7]. Propionic anhydride makes covalent bonds with non-modified or monomethylated lysines and with the N-termini of proteins.
5. Treat gel slices for 1 h at 37 °C with 100 ml of 30% propionic anhydride in methanol and 40 ml of 50 mM NH_4HCO_3 [7], followed by two 10-min washes in 100 mM NH_4HCO_3, one wash in 50% acetonitrile, 100 mM NH_4HCO_3, and one wash in acetonitrile.

6. Dry gel slices and digest at 37 °C overnight using 0.4 mg of sequencing grade trypsin.
7. Acidify the digests in 0.5 % TFA, lyophilize, resuspend in 40 ml of 50 mM NH_4HCO_3, and propionylate again in 100 ml of 30 % propionic anhydride in methanol for 1 h at 37 °C, lyophilized and resuspended in 20 ml of 0.1 % of formic acid. The second propionylation modifies the newly created N-terminal ends after trypsin digestion. These conditions give complete lysine and N-terminal propionylation, but also chemical methylations that can be detected using deuterated methanol (methanol-d4) for the propionic anhydride dilution.
8. Run the obtained peptide mixtures on a Nano C18 PepMap 100 pre-column (5 mm, 100 Å, 300 mm I.D. × 1 mm), coupled with a column of 75 mm I.D. × 15 cm with the same resin (LC Packings). The Nano-flow-High Pressure Liquid Chromatography LC (LC Packings) is directly coupled to an electrospray ionization system on an ion-trap mass spectrometer (ESI/MS-MS) (Thermo Finnigan LCQ Deca XP).
9. Proceed with mass spectrometry analysis to identify complex composition and PTM of histones (*see* **Note 6**).

4 Notes

1. All steps must be performed on ice if not indicated otherwise. Keep all buffers at 4 °C and perform all centrifugation steps at 4 °C throughout the entire procedure. FLAG and HA resins are centrifuged 2 min at 1000 × *g*. Use low binding tubes during all steps.
2. If the NSF will be analyzed, leave the supernatant from step in Subheading 3.2.5 on ice during preparation of chromatin-bound fraction and then treat both fractions simultaneously.
3. Take the NCF (Subheading 3.3) and add 600 µl washed FLAG resin to each experimental point. Do equally for NSF (Subheading 3.2) and CF (Subheading 3.2) if interested.
4. If the signal difference is clear between the cell line specifically overexpressing a FLAG-HA-tagged protein and the control cell line, proceed to step in Subheading 3.6.
5. Products and materials used to optimize mass spec analyses: ddH_2O or Milli-Q; change gloves very often; wear a lab coat all the time, use exclusively pre-cat 4–12 % SDS-PAGE gels to avoid extra-contamination of samples, especially the gradient gels and their buffers; use tubes with low adherence if possible to minimize the loss of material. Other precautions to avoid contamination of mass spec samples are the following: clean the bench, clean the pipettes with alcohol, use new pipette tip boxes, wash the gloves just after wearing them, and wash all

materials to be used (SDS-PAGE system, dounces, boxes, etc.) with detergents (e.g., cleaning solution 7×, ICN- Cat No 76-670-95) in distilled water (more than 10 times).

6. For mass spectrometry, the five most intense ions of the MS scan are subjected to fragmentation (MS-MS) without any data-dependent scan. The interpretation of the mass spectrometry data can be performed with the BioWorks software version 3.2 (Thermo Scientific). For example, for lysine methylation, a bank of peptides from the histones cut at arginine residues can be indexed with permanent add mass for the N-terminus and lysine of 56.025 Da and three modifications: K– 14.015 for acetylation or trimethylation, K+14.015 Da for a monomethylation, and K– 27.995 Da for a dimethylation. This set-up allows automation of analysis of the MS raw data. Each raw dataset can then be analyzed to check for combinations of modifications that might have been missed by the automated method.

Acknowledgments

Work in the Ait-Si-Ali lab was supported by the Association Française contre les Myopathies Téléthon (AFM-Téléthon); Institut National du Cancer (INCa); Agence Nationale de la Recherche (ANR), Fondation Association pour la Recherche sur le Cancer (Fondation ARC); Groupement des Entreprises Françaises pour la Lutte contre le Cancer (GEFLUC); Université Paris Diderot and the "Who Am I?" Laboratory of Excellence #ANR-11- LABX-0071 funded by the French Government through its "Investments for the Future" program operated by the ANR under grant #ANR-11-IDEX-0005-01. S.B. was a recipient of fellowships from Fondation ARC and Mairie de Paris.

References

1. Nakatani Y, Ogryzko V (2003) Immunoaffinity purification of mammalian protein complexes. Methods Enzymol 370:430–444
2. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. doi:10.1126/science.1231143, science.1231143 [pii]
3. Fritsch L, Robin P, Mathieu JR, Souidi M, Hinaux H, Rougeulle C, Harel-Bellan A, Ameyar-Zazoua M, Ait-Si-Ali S (2010) A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol Cell 37(1):46–56. doi:10.1016/j.molcel.2009.12.017, S1097-2765(09)00921-6 [pii]
4. Ouararhni K, Hadj-Slimane R, Ait-Si-Ali S, Robin P, Mietton F, Harel-Bellan A, Dimitrov S, Hamiche A (2006) The histone variant mH2A1.1 interferes with transcription by down-regulating PARP-1 enzymatic activity. Genes Dev 20(23):3324–3336
5. Robin P, Fritsch L, Philipot O, Svinarchuk F, Ait-Si-Ali S (2007) Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. Genome Biol 8(12):R270
6. Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705
7. Bonaldi T, Imhof A, Regula JT (2004) A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. Proteomics 4(5):1382–1396

Chapter 3

Efficient Preparation of High-Complexity ChIP-Seq Profiles from Early *Xenopus* Embryos

George E. Gentsch and James C. Smith

Abstract

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) has become a powerful tool to acquire a precise and genome-wide snapshot of many chromatin features in vivo. These chromatin profiles are obtained by immunoprecipitation of cross-linked chromatin fragments to enrich the feature of interest. Sequencing and aligning the underlying DNA sequences to the genome make it possible to virtually reconstruct the global distribution of most chromatin features. We present here recent improvements to the ChIP-seq protocol by means of *Xenopus* embryos to prepare high-complexity DNA libraries from small amounts of biological material. This approach allows researchers to explore the landscape of chromatin regulators and states in early vertebrate embryos or in any biological entity with small numbers of cells.

Key words Chromatin immunoprecipitation, Next-generation sequencing, ChIP-seq, Library complexity, Post-sequencing analysis, *Xenopus* embryo

1 Introduction

Chromatin—a dynamic and complex assembly of protein, RNA, and DNA—regulates the transcriptional output of the genome in all living organisms. Embryonic development, for example, involving cell differentiation, cell movement, and pattern formation, is regulated by when, where, and how proteins interact with chromatin. The interaction of a specific protein with a single genomic locus in vivo was first revealed in the early 1980s by antibody-mediated enrichment of chromatin (ChIP) followed by diagnostic PCR [1, 2]. Microarray (chip) and next-generation sequencing technologies in the twenty-first century have now replaced PCR, and combined with ChIP these approaches have led to the simultaneous discovery of thousands of genomic loci with specific chromatin features, such as transcription factor binding events or posttranslational histone modifications [3–5]. The most popular approach to create these chromatin profiles is ChIP followed by

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_3,

next-generation sequencing (ChIP-seq). This technique provides considerable advantages over ChIP-PCR and ChIP-chip with respect to comprehensive coverage and positional resolution. Prerequisites for this method are a ChIP-grade antibody, a protocol to efficiently extract and shear chromatin, and a sequenced genome.

This protocol describes a modified ChIP-seq approach for use with early *Xenopus* embryos. It introduces considerable improvements to previously published protocols [6, 7] to allow the genome-wide reconstruction of high-complexity chromatin profiles from as few as 10,000 cells. The same steps are required for profiling the chromatin of any other organism. However, it is likely that the first steps of fixation and chromatin extraction will need some modification due to differences in tissue texture and size.

Briefly, *Xenopus tropicalis* embryos are treated with formaldehyde to cross-link chromatin proteins to nearby genomic DNA. Postfixation embryos can be dissected to select specific anatomical regions if required. Chromatin is extracted, solubilized, and fragmented by sonication. Because fragmentation reduces the amount of DNA associated with any chromatin feature of interest, it introduces the positional resolution required to allocate these features to their genomic loci. Next, ChIP-grade antibodies are used to recognize specific chromatin features and to enrich these by coupling the antibody-chromatin complex to magnetic beads followed by extensive washing. Upon reversal of cross-linking and purification, the co-immunoprecipitated DNA fragments become part of an indexed paired-end ChIP-seq library for next-generation sequencing. This is achieved with a minimal number of PCR cycles to maintain the biological complexity of DNA fragments. Complexity is an estimate of information redundancy as judged by the duplication levels of sequencing reads potentially representing the same ChIP fragment. We consider more than 75% of unique single-end reads from a standard sequencing run yielding around 30 to 40 million reads as high complexity. Post-sequencing consists of aligning the DNA fragments to the reference genome and finding sites of enriched alignment pinpointing the positions of chromatin features in vivo. In doing so, a high complexity is desirable to accurately demarcate chromatin features by correctly discriminating signal from noise.

2 Materials

2.1 Xenopus Embryo Manipulation and Chromatin Cross-Linking

1. 10× MMR: 1 M NaCl, 20 mM KCl, 20 mM $CaCl_2$, 10 mM $MgSO_4$, 50 mM HEPES. Adjust pH to 7.5 and sterilize by autoclaving.
2. 0.01× MMR.

3. 36–38% formaldehyde (stabilized with maximal 15% methanol).
4. Capped glass vial with capacity of 5–10 ml.
5. HEG: 1 mM HEPES, pH 7.5, 1 mM EDTA, 20% glycerol. Store at 4 °C.

2.2 Preparation of Embryo Extracts

1. Chromatin extraction and washing buffer 1 (CEWB1): 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate, 0.1% SDS. Store at 4 °C.
2. 1 mM DTT.
3. Protease inhibitor tablets.
4. 100 mM orthovanadate.
5. 500 mM NaF.

2.3 Chromatin Solubilization and Fragmentation

1. Ultra-sonicator with sound enclosure. If a probe is required, use a tapered microtip (approximately 1/16 in.).
2. Score and clip a 15 mL conical polystyrene tube at the 7 mL mark. Use this tube to contain and sonicate embryo extracts with a microtip.
3. Siliconized (non-stick) low-retention 1.5 mL microcentrifuge tubes.

2.4 Chromatin Immunoprecipitation

1. ChIP-grade antibody.
2. Protein G magnetic beads.
3. Magnetic rack to collect magnetic beads.
4. Washing buffer 2 (WB2): 10 mM Tris–HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate, 0.1% SDS. Store at 4 °C.
5. Washing buffer 3 (WB3): 10 mM Tris–HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% Igepal CA-630, 1% sodium deoxycholate. Store at 4 °C.
6. TEN: 10 mM Tris–HCl pH, 8.0, 150 mM NaCl, 1 mM EDTA. Store at 4 °C.

2.5 Chromatin Reverse Cross-Linking and DNA Purification

1. SDS elution buffer: 50 mM Tris–HCl, pH 8.0, 1 mM EDTA, 1% SDS.
2. 5 M NaCl.
3. Hybridization oven.
4. TE pH 8.0.
5. RNase A (20 μg/μL).
6. Proteinase K (20 μg/μL).

7. PCR purification kit with 1% SDS tolerance (e.g., QIAquick from Qiagen).
8. 5× DNA loading buffer: 0.2% Orange G, 30% glycerol, 60 mM EDTA, pH 8.0.
9. 100 bp and 1 kb DNA ladder.
10. 1.5 mL phase-lock gel heavy tubes.
11. Phenol/chloroform/isoamyl alcohol (25:24:1) pH 7.9.
12. 80% and 100% ethanol.
13. GlycoBlue (15 μg/μL).
14. DNA elution buffer: 10 mM Tris–HCl, pH 8.5.
15. Fluorometer and high sensitivity reagents for detecting double-stranded DNA.

2.6 Preparation of Indexed Paired-End ChIP-Seq Library

1. ChIP-seq library preparation kit includes end repair, A-tailing, DNA ligation, and PCR reagents (KAPA Hyper Prep Kit).
2. Y-adapter (TruSeq) oligonucleotides for Illumina sequencing platforms (HPLC quality):
   *****, phosphorothioate bond
   (P), phosphate group
   italics, region that hybridizes to form the Y-adapter
   (a) Universal (5′ Illumina P5 – read 1 sequencing primer): 5′AATGATACGGCGACCACCGAGATCT –ACACTCT TTCCCTACACGAC*GCTCTTCCGATC******T.
   (b) Indexed (5′ read 2/index sequencing primer – index – Illumina P7): 5′(P)*GATCGGAAGAGC*ACACGTCTGA ACTCCAGTCAC – NNNNNN – ATCTCGTATGCCGTCTTCTGCTT*****G.
   (c) Index 1: ATCACG, index 2: CGATGT, index 3: TTAGGC, index 4: TGACCA, index 5: ACAGTG, index 6: GCCAAT, index 7: CAGATC, index 8: ACTTGA, index 9: GATCAG, index 10: TAGCTT, index 11: GGCTAC, index 12: CTTGTA, index 13: AGTCAA, index 14: AGTTCC, index 15: ATGTCA, index 16: CCGTCC, index 18: GTCCGC, index 19: GTGAAA, index 20: GTGGCC, index 21: GTTTCG, index 22: CGTACG, index 23: GAGTGG, index 25: ACTGAT, index 27: ATTCCT.
   (d) Make Y-adapters by annealing universal and indexed oligonucleotides at 50 μM in 10 mM Tris–HCl, pH 8.0, 0.1 mM EDTA, and 50 mM NaCl. Incubate equimolar mix at 95 °C for 1 min before cooling it down to 4 °C over a period of 30 min. Store Y-adapters as 10 μM stock solution at −20 °C.
3. Paired-end PCR primers (desalted).

   (a) Forward (Illumina P5): 5′ AATGATACGGCGACCA CCGA*G.
   (b) Reverse (Illumina P7): 5′ CAAGCAGAAGACGGCATA CGA*G.
   (c) Make PCR primer mix with both forward and reverse primer at 5 μM.
4. Solid-phase reversible immobilization (SPRI) beads (Agencourt AMPure XP).
5. 96-well low-volume (about 300–400 μL) microplate with V-shaped bases.
6. Adhesive PCR film for microplates.
7. Magnetic stand for 96-well microplates.
8. Plate centrifuge equipped with swing-out microplate buckets.
9. E-gel EX agarose gel, 2%, SYBR Gold, 11 wells (Life Technologies).
10. Gel electrophoresis system for E-gel (Life Technologies).
11. TrackIt 100 bp DNA ladder (Life Technologies).
12. Safe imager viewing glasses (Life Technologies).
13. Gel knife (Life Technologies).
14. Disposable scalpels.
15. QG buffer (Qiagen).
16. Isopropanol.
17. PE wash buffer (Qiagen).
18. MinElute column (Qiagen).
19. Chip-based capillary electrophoresis system (e.g., Bioanalyzer).

2.7 Post-sequencing Analysis

1. Multicore Unix-style computer (at least 8 GB RAM and 500 GB free disk space).
2. Install FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) command-line tool to check the quality and complexity of the sequencing reads.
3. Install samtools [8] to manipulate sequencing files. Create indexed FASTA file (*genome.fa.fai*) from the FASTA genome sequence file at the command line (>):
   `> samtools faidx /path/to/genome.fa.`
4. Install short-read alignment tool Bowtie [9] for mapping reads to the reference genome. Create Bowtie index from the FASTA genome sequence file:
   `> bowtie-build /path/to/bowtie/index/genome.fa xenopus.`

5. Install HOMER [10] to manipulate reads and call regions of enriched alignment and normalize alignment files for visualization. Customize HOMER using FASTA genome sequence file and GTF gene annotation file:
   ```
   > loadGenome.pl -name xenopus -org null
   -fasta /path/to/genome.fa -gtf path/to/
   genes.gtf.
   ```
6. Install Integrative Genome Viewer [11, 12] to visualize chromatin profiles. Upload indexed FASTA file (genome.fa and *genome.fa.fai* file in the same folder) and GTF gene annotation file to create a reference track for the genome viewer.

3 Methods

All *Xenopus* work complies fully with the UK Animals (Scientific Procedures) Act 1986 as implemented by the Francis Crick Institute.

3.1 Xenopus Embryo Manipulation and Chromatin Cross-Linking

1. Follow standard protocols [13] to fertilize *Xenopus* eggs in vitro and to de-jelly and culture embryos.
2. Transfer de-jellied embryos at the desired developmental stage to a capped glass vial (*see* **Note 1**).
3. Wash the embryos briefly once with 0.01× MMR.
4. Refill vial with 0.01× MMR. Move vial to the fume hood and add formaldehyde to a final concentration of 1%. Fix embryos for 15–45 min at room temperature (*see* **Note 2**).
5. Terminate the fixation reaction by briefly rinsing the embryos three times with ice-cold 0.01× MMR (*see* **Note 3**).
6. *Optional*: Dissect embryos in cold 0.01× MMR to isolate the anatomical region of interest.
7. Aliquot embryos into 2 ml microcentrifuge tubes on ice in batches of approximately 250 embryos per tube (*see* **Note 4**).
8. Remove as much 0.01× MMR as possible. Skip the next step if you continue the same day with Subheading 3.2.
9. Equilibrate embryos in 250 µL cold HEG buffer (*see* **Note 5**). Once the embryos have settled to the bottom of the tube, remove as much liquid as possible and snap-freeze in liquid nitrogen. Store at –80 °C.

3.2 Preparation of Embryo Extracts

1. Supplement CEWB1 with protease inhibitor tablet and 0.5 mM DTT. If using phospho-specific antibodies for ChIP, further add orthovanadate and NaF to 0.5 mM and 2.5 mM, respectively (*see* **Note 6**).
2. Keep samples and buffers on ice during the preparation of embryo extracts.

3. Homogenize fixed embryos in ice-cold CEWB1 by pipetting up and down. For upscaling, transfer homogenates to 50 mL centrifuge tubes.
4. Keep on ice for 5 min.
5. Spin homogenates in a refrigerated centrifuge (4 °C) at 1000 × *g* for 5 min. Aspirate the supernatant and any biological material stuck to the wall.
6. Repeat **steps 3–5**.
7. Resuspend pellet in 1–3 mL CEWB1 (*see* **Note 7**).
8. Keep on ice or at 4 °C to proceed with Subheading 3.3 on the same or following day. For later use, snap-freeze in liquid nitrogen and store at –80 °C.

3.3 Chromatin Solubilization and Fragmentation

1. If using a probe-equipped sonicator for chromatin shearing, attach an empty custom-built sonication tube (*see* Subheading 2) to a plastic beaker filled with ice water via a short thermometer clamp.
2. Transfer the embryo extract to the chilled sonication tube (*see* **Note 8**).
3. If necessary place the beaker on a laboratory jack and adjust the height so that the sonicator microtip is submersed in the extract to about two-thirds of the volume depth and centered without touching the tube wall.
4. Sonicate sample at 6–18 W for 4–10 min in total. Microtip-mediated sonication requires breaks of 1 min every 30 s to keep temperature low. Isothermal focused ultra-sonication can be run continuously (*see* **Note 9**).
5. Transfer the extract into prechilled 1.5 mL nonstick tubes and spin at 16,000 × *g* for 5 min at 4 °C.
6. Transfer the supernatant containing solubilized and sheared chromatin to prechilled 1.5 mL nonstick tubes.
7. Collect 50 μL of the supernatant to check whether the chromatin was sheared successfully (*see* **Note 10** and Fig. 1).
8. Use the rest of the supernatant for the input sample and chromatin immunoprecipitation. Samples can be stored at 4 °C for up to 2 days. Snap-freeze samples as aliquots in liquid nitrogen for long-term storage at –80 °C.

3.4 Chromatin Immunoprecipitation

1. Transfer approximately 1% of the chromatin to a new 1.5 mL nonstick tube. This sample will be used later as ChIP input. Keep it at 4 °C until the ChIP samples are due for reverse cross-linking (*see* Subheading 3.5, **step 1**).
2. Add the ChIP-grade antibody to an appropriate amount of chromatin (*see* **Notes 1** and **11**).

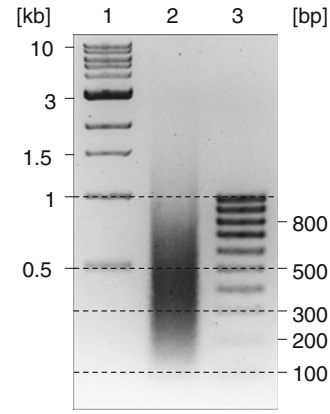

Fig. 1 Validation of sonication-mediated fragmentation of cross-linked chromatin from early gastrula embryos. The image shows the results of size-separating sheared genomic DNA (lane 2) on a 1.8% agarose gel by electrophoresis. Successful sonication results in an asymmetric distribution of DNA fragments ranging from 100 to 1000 bp and peaking between 300 and 500 bp. Here embryos were fixed for 25 min with 1% formaldehyde at room temperature. The chromatin was sheared through 18 cycles (30 s) of sonication with a microtip (12 W) in a volume of 3 ml

3. Incubate overnight on a vertical rotator at 4 °C.
4. Wash an appropriate amount of antibody-compatible magnetic beads once with CEWB1 for 5 min at 4 °C. Use a minimum of 30 µL per ChIP (*see* **Note 12**).
5. Add washed beads to the ChIP sample and incubate for another 4 h on the rotator at 4 °C.
6. Wash beads twice with each washing buffer (CEWB1, WB2 and WB3) for 5 min at 4 °C (*see* **Note 13**).
7. Rinse beads once with TEN buffer.
8. Resuspend beads in 50 µL TEN buffer per tube and transfer them into a new 1.5 mL nonstick tube. Pool here if several tubes for a single ChIP experiment are in use.
9. Collect beads at the bottom of tube by means of the magnetic rack and centrifugation at 1000 × *g* (4 °C).
10. Discard as much liquid as possible without disturbing the pellet of beads.
11. Elute the immunoprecipitated chromatin by resuspending the beads in 100 µL SDS elution buffer. Vortex and incubate for 15 min at 65 °C before spinning the beads at 16,000 × *g* for 30 s. Transfer the ChIP eluate to a new 1.5 mL nonstick tube.
12. Repeat the last step and combine ChIP eluates.

3.5 Chromatin Reverse Cross-Linking and DNA Purification

1. Adjust input sample with SDS elution buffer to 200 µL.
2. Add 10 µL 5 M NaCl to both ChIP and input samples.
3. Reverse cross-link chromatin by incubating samples for 6–15 h (overnight) at 65 °C in a hybridization oven.
4. Add 200 µL TE buffer and 40 µg RNase A. Incubate for 1 h at 37 °C.
5. Add 40 µg proteinase K. Incubate for 3 h at 55 °C.
6. Pre-spin 1.5 mL phase-lock gel heavy tubes at 16,000 × *g* for 30 s at room temperature.
7. Transfer samples to pre-spun phase-lock tubes.
8. Add 400 µL phenol/chloroform/isoamyl alcohol (25:24:1) with pH 7.9.
9. Spin tubes at 16,000 × *g* for 5 min at room temperature.
10. Transfer upper phase to a new 1.5 mL nonstick tube.
11. Add 16 µL 5 M NaCl, 800 µL 100% ethanol, and 15 µg GlycoBlue.
12. Mix by inverting tube four to six times before storing samples overnight at −20 °C.
13. Spin tubes at more than 16,000 × *g* for 1 h at 4 °C.
14. Carefully discard supernatant without disturbing the blue DNA pellet.
15. Add 500 µL 80% ethanol and spin at more than 16,000 × *g* for 2 min at 4 °C.
16. Air-dry DNA pellet for 10 min at room temperature.
17. Add 11 µL DNA elution buffer to dry DNA pellet.
18. Leave samples on ice for 30 min to ensure that the DNA is completely dissolved.
19. Determine the concentration of 1 µL ChIP sample using a fluorometer and high sensitivity reagents to detect double-stranded DNA. Dilute the input sample so it falls within the detection range of the fluorometer (*see* **Notes 1** and **14**).

3.6 Preparation of Indexed Paired-End ChIP-Seq Library

1. Set up DNA end repair and A-tailing for both ChIP and input DNA in separate PCR tubes as follows: 10 µL DNA (500 pg to 5 ng), 40 µL molecular-grade water, 7 µL end repair and A-tailing buffer, and 3 µL end repair and A-tailing enzyme mix.
2. Use PCR machine with a heated lid (98–105 °C) to run reaction for 30 min at 20 °C, followed by 30 min at 65 °C before cooling to 4 °C.
3. Transfer 190 µL SPRI beads per library to 1.5 mL tube for equilibration to room temperature.

4. Set up adapter ligation by adding the following reagents to the previous reactions: 5 μL molecular-grade water, 30 μL ligation buffer, 10 μL DNA ligase, and 5 μL Y-adapters (*see* **Note 15** for Y-adapter concentrations). Mix well as the ligation buffer is quite viscous.
5. Incubate for 20 min at 20 °C.
6. Add 88 μL SPRI beads, mix well, and transfer the bead suspension to a 96-well microplate.
7. Wait 5 min before transferring the plate to the magnetic stand.
8. Wait 3–5 min until the beads have separated from the supernatant.
9. Discard the supernatant. Add 180 μL 80% ethanol without disturbing the beads.
10. Wait 1 min before repeating **step 9**.
11. Discard the supernatant and seal the plate with an adhesive PCR film.
12. Remove the plate from the magnetic stand and spin it at 200 × *g* for 1 min.
13. Put the plate back onto the magnetic stand and remove the remaining supernatant with a 20 μL pipette tip.
14. Air-dry until the bead pellets show multiple cracks. Avoid any draft as the dry beads dislodge easily.
15. Add 21.5 μL DNA elution buffer to the dried bead pellets.
16. Remove the plate from the magnetic stand and resuspend the beads well.
17. Wait 5 min before putting the plate back onto the magnetic stand.
18. After 30 s transfer 20 μL eluate to a new PCR tube.
19. Set up PCR to make adapter-ligated DNA fragments double stranded (*see* **Note 16**). Add 5 μL paired-end primer mix and 25 μL KAPA high-fidelity polymerase master mix.
20. Run PCR with a heated lid (98–105 °C) as follows: 45 s at 98 °C followed by 5 cycles of 15 s at 98 °C, 30 s at 60 °C and 30 s at 72 °C, followed by 1 min at 72 °C before cooling to 4 °C.
21. Add 50 μL SPRI beads to each PCR reaction, mix well, and transfer the bead suspension to a 96-well microplate.
22. Repeat **steps 7–14**.
23. Add 20 μL DNA elution buffer to the dried bead pellets.
24. Remove the plate from the magnetic stand and resuspend the beads well.
25. Wait 5 min before putting the plate back onto the magnetic stand.

26. Load a 2% E-gel EX agarose gel onto the E-gel electrophoresis system.
27. Load the DNA eluates into separate wells of the E-gel with at least one empty lane between each sample and the DNA ladder.
28. Use 20 µL 1:10 dilution of TrackIt 100 bp DNA ladder and fill all remaining wells with 20 µL DNA elution buffer.
29. Run the gel for 10 min.
30. Open the gel cassette with a gel knife.
31. Place the open gel cassette onto the blue light transilluminator.
32. Wear safe imager viewing glasses to cut gel slices for each library containing DNA ranging from 250 to 450 bp in size with a new disposable scalpel.
33. Transfer the gel slices to new 1.5 mL nonstick tubes.
34. Add 350 µL QG buffer to each tube (*see* **Note 17**).
35. Shake the tubes at room temperature until the gel has completely dissolved.
36. Add 70 µL isopropanol and mix.
37. Apply mix to MinElute columns and spin at 9000 × *g* for 30 s.
38. Collect the flow-through and apply it again to the same column.
39. Spin at 9000 × *g* for 30 s. Discard the flow-through.
40. Add 500 µL QG buffer to the column. Spin at 16,000 × *g* for 30 s.
41. Discard the flow-through. Add 750 µL PE wash buffer to the column.
42. Wait 2 min before spinning the column at 16,000 × *g* for 30 s.
43. Discard the flow-through and dry the column by spinning it at 16,000 × *g* for 2 min.
44. Place the column into a new 1.5 mL nonstick tube.
45. Add 11 µL DNA elution buffer to the column. Wait 2 min before spinning the column at 16,000 × *g* for 30 s.
46. Repeat the last step.
47. Transfer 20 µL eluate to a new PCR tubes.
48. Set up PCR by adding 5 µL paired-end primer mix and 25 µL KAPA high-fidelity polymerase master mix.
49. Run PCR with a heated lid (98–105 °C) as follows: 45 s at 98 °C followed by 3–13 cycles (*see* **Note 18** for the approximate number of PCR cycles) of 15 s at 98 °C, 30 s at 60 °C, and 30 s at 72 °C, followed by 1 min at 72 °C before cooling to 4 °C.
50. Repeat **step 21** followed by **steps 7–14**.

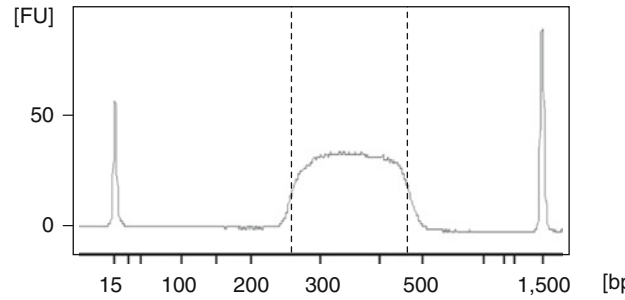

Fig. 2 Preparation of an indexed paired-end ChIP-seq library containing size-selected DNA fragments. The electropherogram shows the DNA fragment distribution of a quality control approved library created from 1 ng co-immunoprecipitated DNA (12 PCR cycles). It shows the expected range of DNA templates of 250–450 bp and no adapter dimers, which would be detected at 120 bp. These templates contain the DNA fragment of interest (about 130–320 bp), which is flanked by the universal (58 bp) and the indexed adapter (64 bp)

51. Add 14 μL DNA elution buffer to the dried bead pellets.
52. Remove the plate from the magnetic stand and resuspend the beads well.
53. Wait 5 min before putting the plate back onto the magnetic stand.
54. Transfer 12 μL eluate to a new 1.5 mL nonstick tube.
55. Determine the concentration of 1 μL library using a fluorometer and high sensitivity reagents to detect double-stranded DNA.
56. Determine library integrity with a chip-based capillary electrophoresis system (*see* **Note 19** and Fig. 2).
57. If multiplexing is desired, mix quality control approved libraries at equimolar ratios (*see* **Note 20**).
58. Run libraries on an Illumina sequencing platform.

3.7 Post-sequencing Analysis

1. Concatenate the gzip-compressed FASTQ files if the sequencing results of a single library are split into smaller data packages (*see* **Note 21**).
   ```
   > cat /path/to/*.fastq.gz>ChIP.fastq.gz.
   ```
2. Check the quality of the sequencing data with FastQC (*see* **Note 22** and Fig. 3).
   ```
   > fastqc ChIP.fastq.gz.
   ```
3. Decompress the FASTQ file and trim reads to remove any potential adapter contamination with *homerTools trim*. Allow one mismatch and discard any processed reads shorter than 28 bases.
   ```
   > gzip -cd ChIP.fastq.gz>ChIP.fastq
   > homerTools trim -3 GATCGGAAGAGCACACGTCT
   -mis 1 -min 28 ChIP.fastq.
   ```

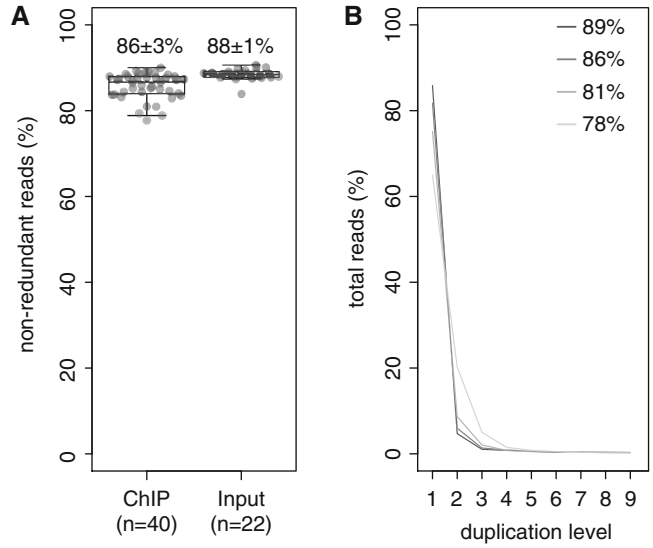

Fig. 3 Library complexity based on the counts of nonredundant reads from single-end sequencing. (**a**) This *boxplot* shows the percentage of nonredundant reads calculated from 62 ChIP-seq libraries (40 ChIPs and 22 inputs). (**b**) This plot shows the direct correlation of duplication levels of four selected ChIP-seq libraries and library complexity (78, 81, 86, and 89 %). All these libraries are of high complexity (more than 75 %). They are made from 0.5 to 15 ng of co-immunoprecipitated DNA with 7–13 PCR cycles in total

4. Align processed single-end reads to the reference genome using Bowtie for which an index (*xenopus*) has been generated (*see* Subheading 2). Only keep uniquely mapped reads (*-m 1*) and limit the number of mismatches within the first 28 bases (seed) to 1 (*-n 1*). Only report best alignments (*--best*) in terms of number of mismatches in the seed (*--strata*) and quality values at the mismatched position. Save alignment in SAM format (*-S*). Set the number of threads to the number of available computer cores and increase the amount of chunk memory per thread to 256 megabytes (*see* **Note 23**).
```
> export BOWTIE_INDEXES=/path/to/bowtie/index
> bowtie -m 1 -n 1 --best --strata -p [# threads] -S --chunkmbs 256 xenopus ChIP.fastq.trimmed>ChIP.sam.
```
5. Use two HOMER commands to convert the SAM file into a virtual chromatin profile in indexed binary format (*-bigWig*) normalized to ten million mapped reads (*-norm 1e7*). Remove any redundant reads by reducing the number of tags per base pair to one (*-tbp 1*, *see* **Note 24**). The conversion also requires an indexed FASTA file (*genome.fa.fai*) created from the reference genome (*see* Subheading 2).
```
> makeTagDirectory ChIP/ -single -tbp 1 ChIP.sam
```

```
> makeUCSCfile ChIP/ -bigWig /path/to/genome.
fa.fai -fsize 1e20 -norm 1e7 -o ChIP.bw.
```

6. Use HOMER *findpeaks* to find sites of enriched read alignment when comparing the ChIP and input tag directories created by *makeTagDirectory* (*see* previous step). Below an example is shown for finding peaks (*-style factor*, *see* **Note 25**) that are at a minimal distance of 250 bp from each other (*-minDist 250*). Autocorrelation (calculating the distance between 5′ positions of plus and minus strand alignment at peaks) is used to estimate the fragment size of the ChIP library, while reads of the input library are extended to an average fragment length of 175 bp. The false discovery rate (FDR) of peak detection is set to 0.1% (*-fdr 0.001*). In addition, peaks need a minimal threefold read enrichment (ChIP versus input, *-F 3*) and a lower clonality value than 0.97 (expected number of unique positions versus the total number of reads within peak, *-C 0.97*). The local density filter is disabled (*-L 0*). The mappable genome size of *X. tropicalis* is about 1.43 billion bp (*-gsize 1.43e9*). Subsequently, the Perl script *pos2bed.pl* converts the format of the peak file to BED, which can be displayed together with the bigWig file in the Integrative Genome Viewer.

```
> findPeaks ChIP/ -style factor -i Input/
-inputFragLength 175 -minDist 250 -fdr 0.001
-F 3 -C 0.97 -L 0 -gsize 1.43e9>ChIP_peaks.
txt
> pos2bed.pl ChIP_peaks.txt -float ChIP_
peaks.bed.
```

7. Explore the chromatin profile and the peak calling by loading the bigWig (*ChIP.bw*, *Input.bw*) and the BED file (*ChIP_peaks.bed*) to the Integrative Genome Viewer (*see* Fig. 4).

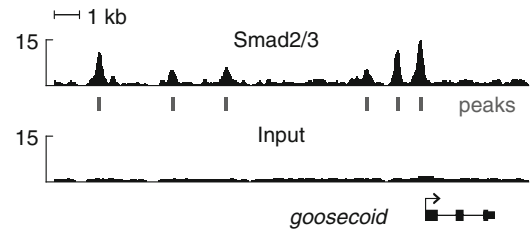

Fig. 4 An excerpt of the genome-wide Smad2/3 chromatin profile around the homeobox gene *goosecoid* from *X. tropicalis* embryos at early gastrula stage. Both profiles, Smad2/3 ChIP and input, are normalized to ten million uniquely mapped and nonredundant reads. The peaks of the Smad2/3 ChIP are called by HOMER *findpeaks* with a false discovery rate (FDR) of 0.1% as outlined in **step 6** of Subheading 3.7. The profiles are normalized pileups of reads extended to an average DNA fragment length. The peaks represent the genomic positions of Smad2/3 occupancy in vivo

Use programming platforms R/Bioconductor or MATLAB to further manipulate and visualize ChIP-seq data.

4 Notes

1. The number of *X. tropicalis* embryos required for a single ChIP-seq experiment depends on the chosen developmental stage and several properties of the protein of interest, in particular, its expression pattern, its nuclear concentration, and its chromatin distribution. In addition, its cross-linking propensity and its epitope accessibility may play a role, but these characteristics are difficult to predict, let alone change in vivo. The numbers are best determined empirically to achieve yields of 500 pg or more of co-immunoprecipitated DNA. Libraries can be generated from far less than 500 pg, albeit at the expense of complexity (*see* **Note 14**). We have successfully created high-complexity libraries from as few as 10,000 cells.
2. Approach the optimal fixation time empirically for each epitope by ChIP followed by quantitative PCR (qPCR) aiming for a high signal-to-noise ratio (yield ratio of positive versus negative control loci). As an antibody control, use normal serum of the same isotype and host animal species as the ChIP-grade antibody. Consider longer fixation times if the protein of interest is expected to have weak (or indirect) DNA binding properties. In addition, early embryos with bigger cells and higher levels of yolk require longer fixation times than those of later stages. However, refrain from fixing *X. tropicalis* embryos for longer than 45 min, as efficient chromatin shearing becomes difficult (*see* **Note 9**).
3. Do not quench formaldehyde with glycine because the adduct (Schiff base) may further react with N-terminal amino-groups or arginine residues [14], rendering chromatin extraction from yolk-rich embryos challenging. Prevent the fixed embryos from making contact with the liquid surface as its tension causes them to burst.
4. Batches of 250 embryos before hatching occupy approximately 250 μL (*X. tropicalis*) or 600 μL (*X. laevis*) in a 2 mL microcentrifuge tube with round bottom.
5. Glycerol of the HEG buffer stabilizes proteins and facilitates both quick thawing on ice and homogenization to make embryo extract.
6. Approximately 8 mL ice-cold CEWB1 in total (4 mL per homogenization, **step 3**) is required for preparing extracts from 250 *X. tropicalis* or 100 *X. laevis* embryos. Keep aside an additional 1–3 mL ice-cold CEWB1 for resuspending final extracts for sonication.

7. The extracts from embryos, especially those of later stages with less yolk, can become fairly transparent. Extracts from early embryos may still contain a substantial amount of yolk; however this does not impede efficient chromatin shearing. If the embryo extract is very viscous, dilute it with CEWB1 to 2 or 3 ml.
8. If using isothermal focused ultra-sonication (Covaris) for chromatin shearing, transfer embryo extracts to the designated container (e.g., 1 mL milliTUBE) and place the container into the appropriate holder.
9. Cross-linked chromatin is solubilized and sheared by sonication, whose settings need to be optimized empirically. The power and time required to efficiently shear chromatin depends on the volume and concentration of the embryo extract as well as the degree of chromatin cross-linking. The following Covaris settings achieve about 11 W and are a good starting point for shearing chromatin in a volume of 1 mL: duty cycle, 5%; intensity, 4; cycles per burst, 200; processing time, 240 s. Pause immediately if the sample begins to froth. Wait until the froth has completely disappeared. If necessary, reposition tube and reduce power before resuming sonication.
10. To visualize the degree of chromatin fragmentation as outlined here and shown in Fig. 1, the sample should contain chromatin derived from more than 100,000 nuclei. Use more sensitive equipment such as E-gel or chip-based capillary electrophoresis to detect lower amounts of DNA. Add 50 µL SDS elution buffer, 4 µL 5 M NaCl, and 20 µg proteinase K to the sample. Incubate for 6–15 h (overnight) at 65 °C preferentially in a hybridization oven to avoid condensation at the bottom of the lid. Purify DNA using a PCR purification kit that tolerates 1% SDS (e.g., QIAquick from Qiagen). If necessary, use 3 M sodium acetate (pH 5.2) to adjust the pH as recommended by the manufacturer. Elute the DNA twice with 11 µL DNA elution buffer. Spike DNA with 8 µg RNase A before adding 5 µL 5× DNA loading buffer. Run entire sample alongside a 100 bp and a 1 kb DNA ladder on a 1.4% agarose gel by electrophoresis. For optimal results, stain gel with a safe nucleic acid staining solution after electrophoresis. Sheared chromatin should show an asymmetric distribution of DNA fragments mainly ranging from 100 to 1000 bp and peaking between 300 and 500 bp.
11. The success of a ChIP experiment depends largely on the quality of the antibody. It is paramount that appropriate controls are conducted to prove its specificity for the epitope of interest [15]. Consider introducing epitope-tagged fusion proteins if ChIP-grade antibodies are not available or if the protein of interest is expressed at very low levels. Such experiments can be informative, because these proteins can be recruited to endogenous binding sites in the right developmental context [16].

The amount of antibody required per ChIP can vary considerably. Normally, 1 μg antibody per million cells expressing the epitope of interest should yield enough co-immunoprecipitated DNA. To more accurately estimate the amount of antibody, run the same ChIP with various amounts of antibody (e.g., 0.2, 1 and 2.5 μg) by ChIP-qPCR (*see* **Note 2**).

12. Read the manufacturer's specification for the antibody binding capacity of the magnetic beads. Usually 5–20 μL beads bind about 1 μg IgG antibody. Protein G magnetic beads are versatile in that they are compatible with many IgG antibodies of different host species.
13. Leave the tubes for at least 30 s in the magnetic rack to avoid any bead loss before discarding the supernatant.
14. ChIP-seq libraries can be created from as little as 10 pg DNA. This corresponds to about 30 million DNA fragments (300 bp on average), of which probably only 10% will be incorporated into a library. The main sources of loss originate from failing adapter ligation and size selection. That means a sequencing run yielding 30 million reads will have a high incidence of the same fragment being sequenced twice or more. This results in a low-complexity profile with poor coverage and depth of sequencing reads, which causes problems in correctly distinguishing signal from noise.
15. To avoid the generation of dimers, adjust the Y-adapter concentrations to the amount of DNA in the reaction as follows: 5 μM Y-adapter (15 ng DNA), 1.5 μM (5 ng), 300 nM (1 ng), 60 nM (200 pg), 12 nM (40 pg).
16. Y-adapters have floppy single-stranded ends, which can cause successfully ligated DNA fragments to migrate inappropriately in gel electrophoresis. Thus, making these DNA fragments completely double-stranded is important to obtain compact libraries of the expected size range.
17. Use 5 volumes (w/v) QG buffer per gel slice. The weight of the gel slice is usually between 50 and 70 mg, which requires the use of 250–350 μL QG buffer to dissolve the agarose. Do not heat chaotropic QG buffer to 50 °C, which can denature A/T rich DNA fragments. Subsequently, spin columns can introduce a considerable GC bias by not absorbing single-stranded DNA very efficiently [17].
18. Minimize the cycle number of the second PCR to maintain the biological complexity of DNA fragments as follows: 3–4 cycles (15 ng DNA), 5–6 (5 ng), 7–8 (1 ng), 9–10 (200 pg), 11–12 (40 pg).
19. Successful ChIP-seq library preparation yields 20–400 ng of double-stranded DNA of the expected size range (250–450 bp) without any adapter dimer contamination (band

around 120 bp) as shown in Fig. 2. Accurate quantification and estimation of the DNA size distribution are important to calculate the correct molarities and to achieve optimal next-generation sequencing (clustering) results. Expect poor library complexities if the amount of starting material is low (*see* **Note 14**) or if the expected yield has not been reached even after 18 PCR cycles in total.

20. Multiplex so that the sequencing of each library yields about 20–30 million single-end reads of at least 36 bp, which is enough to cover the *X. tropicalis* genome with sufficient depth. Consider increasing the number of reads if the chromatin feature of interest is expected to show broad distributions. If high mappability within repetitive regions is desired, increase read length and sequence both ends (paired-end) of the DNA template.
21. The standard format of next-generation sequencing is FASTQ. FASTQ files contain the reads from a single library (de-multiplexed) and the corresponding quality scores for each base call.
22. The successful sequencing of a high-complexity ChIP-seq library should pass all tests. Failures originate mainly from poor sequencing runs, low library complexity, or adapter contamination. Some level of duplication is expected due to the presence of *bona fide* biological duplicates. In addition, because sequencing at one end of the DNA fragment (single-end) provides enough information for most ChIP-seq experiments, it leaves open the possibility that identical reads originate from DNA fragments with a different opposite end.
23. Expect about 50–70 % of single-end reads of 36 bp to map uniquely to the *X. tropicalis* genome assembly of version 7.1 with a maximum of one mismatch within the first 28 bp.
24. The removal of redundant reads eliminates any potential PCR amplification bias. This measure of precaution hardly affects the sensitivity of detecting enriched read alignments [18]. Consider replacing the adapter index with a random bar code to distinguish technical from biological duplicates. However, the gain of random bar coding is often marginal as we routinely reach an average of 87 % nonredundant reads from sequencing ChIP-seq libraries (*see* Fig. 3).
25. Consider generating a blacklist of false-positive peaks caused by the incorrect genome assembly collapsing repetitive sequences into a single copy [19]. Screen the input for peaks and remove these and poorly annotated scaffolds from the list of ChIP peaks with *bedtools intersect* [20].

Acknowledgments

We are grateful for the help from Abdul Sesay and Leena Bhaw from the Advanced Sequencing Facility at the Francis Crick Institute. We thank Thomas Spruce for his comments on the manuscript. G.E.G. and J.C.S. were supported by the Medical Research Council (program number U117597140) and are now supported by the Francis Crick Institute.

References

1. Gilmour DS, Lis JT (1984) Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci U S A 81(14):4275–4279
2. Gilmour DS, Lis JT (1985) In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster. Mol Cell Biol 5(8):2009–2018
3. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. Science 290(5500):2306–2309. doi:10.1126/science.290.5500.2306
4. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2):315–326. doi:10.1016/j.cell.2006.02.041, S0092-8674(06)00380-1 [pii]
5. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830):1497–1502. doi:10.1126/science.1141319, 1141319 [pii]
6. Gentsch GE, Smith JC (2014) Investigating physical chromatin associations across the Xenopus genome by chromatin immunoprecipitation. Cold Spring Harb Protoc 2014 (5). doi:10.1101/pdb.prot080614, 2014/5/pdb.prot080614 [pii]
7. Gentsch GE, Patrushev I, Smith JC (2015) Genome-wide snapshot of chromatin regulators and states in Xenopus embryos by ChIP-Seq. J Vis Exp (96). doi:10.3791/52535
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. doi:10.1093/bioinformatics/btp352, btp352 [pii]
9. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. doi:10.1186/gb-2009-10-3-r25, gb-2009-10-3-r25 [pii]
10. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38(4):576–589. doi:10.1016/j.molcel.2010.05.004, S1097-2765(10)00366-7 [pii]
11. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26. doi:10.1038/nbt.1754, nbt.1754 [pii]
12. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14(2):178–192. doi:10.1093/bib/bbs017, bbs017 [pii]
13. Sive H, Grainger R, Harland R (2000) Early development of Xenopus laevis: a laboratory manual. Cold Spring Harbor Laboratory Press, New York
14. Metz B, Kersten GF, Hoogerhout P, Brugghe HF, Timmermans HA, de Jong A, Meiring H, ten Hove J, Hennink WE, Crommelin DJ, Jiskoot W (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. J Biol Chem 279(8):6235–6243. doi:10.1074/jbc.M310752200, M310752200 [pii]
15. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L,

Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22(9):1813–1831. doi:10.1101/gr.136184.111, 22/9/1813 [pii]

16. Mazzoni EO, Mahony S, Iacovino M, Morrison CA, Mountoufaris G, Closser M, Whyte WA, Young RA, Kyba M, Gifford DK, Wichterle H (2011) Embryonic stem cell-based mapping of developmental transcriptional programs. Nat Methods 8(12):1056–1058. doi:10.1038/nmeth.1775, nmeth.1775 [pii]
17. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ (2008) A large genome center's improvements to the Illumina sequencing system. Nat Methods 5(12):1005–1010. doi:10.1038/nmeth.1270, nmeth.1270 [pii]
18. Chen Y, Negre N, Li Q, Mieczkowska JO, Slattery M, Liu T, Zhang Y, Kim TK, He HH, Zieba J, Ruan Y, Bickel PJ, Myers RM, Wold BJ, White KP, Lieb JD, Liu XS (2012) Systematic evaluation of factors influencing ChIP-seq fidelity. Nat Methods 9(6):609–614. doi:10.1038/nmeth.1985, nmeth.1985 [pii]
19. Pickrell JK, Gaffney DJ, Gilad Y, Pritchard JK (2011) False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. Bioinformatics 27(15):2144–2146. doi:10.1093/bioinformatics/btr354, btr354 [pii]
20. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. doi:10.1093/bioinformatics/btq033, btq033 [pii]

Chapter 4

Systematic Discovery of Chromatin-Bound Protein Complexes from ChIP-seq Datasets

Eugenia Giannopoulou and Olivier Elemento

Abstract

Chromatin immunoprecipitation followed by sequencing is an invaluable assay for identifying the genomic binding sites of transcription factors. However, transcription factors rarely bind chromatin alone but often bind together with other cofactors, forming protein complexes. Here, we describe a computational method that integrates multiple ChIP-seq and RNA-seq datasets to discover protein complexes and determine their role as activators or repressors. This chapter outlines a detailed computational pipeline for discovering and predicting binding partners from ChIP-seq data and inferring their role in regulating gene expression. This work aims at developing hypotheses about gene regulation via binding partners and deciphering the combinatorial nature of DNA-binding proteins.

Key words Combinatorial transcription factor binding, Protein complexes, ENCODE datasets, Protein-protein interactions, ChIP-seq, RNA-seq

1 Introduction

Cis-regulatory elements (CREs), also known as cis-regulatory regions or modules [1, 2], are genomic regions where transcription factors (TFs) bind and affect transcription regulation of nearby genes. Frequently, TFs bind chromatin synergistically, as sets of co-associated transcription factors, and form protein complexes, such as the already known and well-studied AP-1 complex and the Polycomb Repressive Complex 2 (PRC2). The systematic discovery of protein complexes from experimental TF binding data is a computational challenge and has been addressed by several studies published in the past few years [3–13]. These studies focus on the combinatorial TF binding on CREs across different cellular conditions, using datasets from chromatin immunoprecipitation followed by sequencing (ChIP-seq) experiments; ChIP-seq is a genome-wide binding assay that has been invaluable for the identification of TF binding sites and histone modifications (HMs).

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_4,

In this chapter, we describe a computational methodology based on nonnegative matrix factorization (NMF) and regression analysis that: (1) integrates ENCODE ChIP-seq datasets and identifies CREs, (2) discovers potential protein complexes, and (3) predicts their regulatory role and impact on gene expression (Fig. 1). We present the application of this method to a large collection of TF binding data in the H1 human embryonic stem cells (H1 Esc) from the Encyclopedia of DNA elements (ENCODE) project, as shown in our previously published work [5].

NMF is a powerful matrix decomposition and dimensionality reduction technique used to discover patterns and relevant correlations in multidimensional data [14]. Unlike other clustering algorithms (e.g., hierarchical/k-means clustering), NMF provides soft clustering (also known as fuzzy clustering), in which individual data points can belong to several clusters with varying degrees. In the context of transcriptional regulation, NMF's soft clustering is essential because it allows for a TF to belong to multiple complexes and a CRE to be binding site for multiple TFs. NMF has also been used in several biological applications because of its nonnegativity constraint. This provides an intuitive and biologically interpretable decomposition of a multivariate dataset and a natural way to cluster biological data. This is unlike principal components analysis, where eigenvectors with negative sign loadings can be hard to interpret in the context of positively valued variables, such as ChIP-seq read counts.

In the methodology described here, each NMF cluster represents a positive linear combination of the original ChIP-seq read count variables. Consequently, every cluster reveals a binding pattern that represents a set of TFs concurrently found by ChIP-seq at the same CRE. Thus, NMF clusters provide evidence for the existence of potential complexes with one or more TFs. Importantly, protein complex scores, which quantify the presence of each complex in the CRE, characterize each CRE.

To model the effect of each protein complex on a gene, CREs are first associated with the closest transcription start sites (TSS). For each TSS, the weighted complex scores for all associated CREs are summed up to define a Binding Influence Score (BIS) between a complex and a gene. These BIS values are then used as explanatory variables (predictors) in a linear regression model to assess the contribution of a detected protein complex to the absolute mRNA expression value of a gene (response). Finally, the regression model coefficients are used to explain the role of the protein complexes in gene expression. For example, a significant and positive coefficient indicates that the corresponding protein complex positively contributes to mRNA expression values, while a negative coefficient indicates negative (i.e., repressive) contribution.

Overall, this methodology serves as a valuable resource for understanding the collective function and role of cis-regulatory elements and the potential chromatin-bound protein complexes

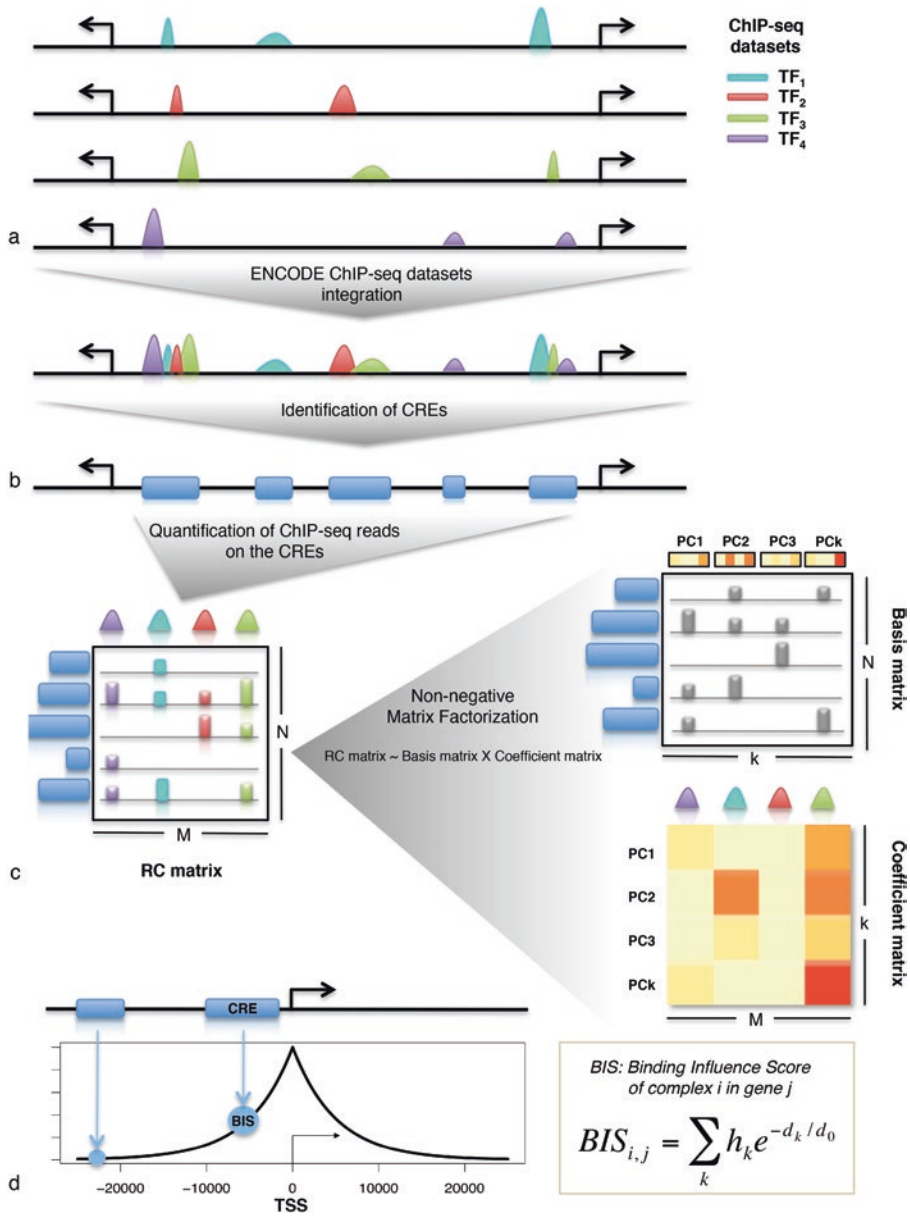

Fig. 1 Modeling gene expression from combinatorial binding. Overview of the methodological workflow. Collection (**a**) and integration of ENCODE ChIP-seq datasets (**b**) into CREs. Quantification of normalized ChIP-seq reads on the CREs (RC matrix) (**c**). NMF analysis applied on the normalized ChIP-seq reads matrix (**d**). NMF decomposes the RC matrix into the basis matrix and the mixture coefficient matrix. Estimation of the CREs that occur within a fixed-range window around a TSS (**e**). Complex scores and the proximity of the CREs to the TSS of a gene are integrated into a Binding Influence Score (BIS) between a protein complex and a gene. These BIS values are used as predictors to assess the contribution of protein complexes to gene expression in the linear regression model

that bind them. The proposed computational approach can also generate hypotheses involving chromatin organization and gene regulation via co-regulators and binding partners.

2 Materials

2.1 Computational Tools

It is recommended the described analysis be performed in a Unix-based operating system (OS), such as Linux, Mac OS X, and more.

1. ChIPseeqer version 2.2 [15].
   Available at: https://gitlab.com/egiannopoulou/ChIPseeqer-2.2/
2. R version 2.15.2 (or later).
   Available at: https://www.r-project.org/
3. NMF version 0.20.6 (R package).
   Available at: https://cran.r-project.org/web/packages/NMF/index.html

2.2 ENCODE Datasets

Although the methodology described here is applicable to any set of ChIP-seq datasets (coming from the same cell type), we use a collection of 31 TFs and 10 HMs ChIP-seq experiments, all in the H1 Esc human cell type, produced under the ENCODE project. Table 1 lists these datasets, their corresponding ENCODE aligned reads files, and URL addresses where the files are available for the reader to download.

1. Cell line: H1 human embryonic stem cells (H1 Esc).
2. ChIP-seq files: hg19 aligned read files (bam format) for each ChIP-seq dataset (Table 1).
3. RNA-seq file: Available in GEO repository, under accession number GSM758566
   ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM758nnn/GSM758566/suppl/GSM758566_hg19_wgEncodeCshlLongRnaSeqH1hescCellPapTranscriptGencV7.gtf.gz

3 Methods

3.1 Preparation of ChIP-seq Datasets

1. Download the aligned reads files (bam format) for each ChIP-seq dataset from the ENCODE Data Coordination Center links in Table 1. There are two replicates for each experiment, corresponding to two bam files per TF/HM dataset.
2. Save both bam files in a separate directory for each ChIP-seq dataset (Fig. 2a). Each TF/HM directory should contain the two bam files; for example, directory ATF should contain

Table 1
List of ENCODE H1 Esc ChIP-seq datasets

ChIP-seq experiment name	Download aligned reads file from ENCODE	
ATF2	wgEncodeHaibTfbsH1hescAtf2sc81188V0422111AlnRep1.bam wgEncodeHaibTfbsH1hescAtf2sc81188V0422111AlnRep2.bam	Link 1
ATF3	wgEncodeHaibTfbsH1hescAtf3V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescAtf3V0416102AlnRep2.bam	Link 1
BCL11A	wgEncodeHaibTfbsH1hescBcl11aPcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescBcl11aV0416102AlnRep2.bam	Link 1
CHD1	wgEncodeBroadHistoneH1hescChd1a301218aStdAlnRep1.bam wgEncodeBroadHistoneH1hescChd1a301218aStdAlnRep2.bam	Link 2
CTCF	wgEncodeHaibTfbsH1hescCtcfsc5916V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescCtcfsc5916V0416102AlnRep2.bam wgEncodeBroadHistoneH1hescCtcfStdAlnRep1.bam wgEncodeBroadHistoneH1hescCtcfStdAlnRep2.bam	Link 1 Link 2
EGR1	wgEncodeHaibTfbsH1hescEgr1V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescEgr1V0416102AlnRep2.bam	Link 1
EP300	wgEncodeHaibTfbsH1hescP300V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescP300V0416102AlnRep2.bam	Link 1
FOSL1	wgEncodeHaibTfbsH1hescFosl1sc183V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescFosl1sc183V0416102AlnRep2.bam	Link 1
GABP	wgEncodeHaibTfbsH1hescGabpPcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescGabpPcr1xAlnRep2.bam	Link 1
HDAC2	wgEncodeHaibTfbsH1hescHdac2sc6296V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescHdac2sc6296V0416102AlnRep2.bam	Link 1
H2AZ	wgEncodeBroadHistoneH1hescH2azStdAlnRep1.bam wgEncodeBroadHistoneH1hescH2azStdAlnRep2.bam	Link 2
JUND	wgEncodeHaibTfbsH1hescJundV0416102AlnRep1.bam wgEncodeHaibTfbsH1hescJundV0416102AlnRep2.bam	Link 1
KDM5A (JARID1A)	wgEncodeBroadHistoneH1hescJarid1aab26049StdAlnRep1.bam wgEncodeBroadHistoneH1hescJarid1aab26049StdAlnRep2.bam	Link 2
MYC	wgEncodeOpenChromChipH1hescCmycAlnRep1.bam wgEncodeOpenChromChipH1hescCmycAlnRep2.bam	Link 3
NANOG	wgEncodeHaibTfbsH1hescNanogsc33759V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescNanogsc33759V0416102AlnRep2.bam	Link 1
POU5F1	wgEncodeHaibTfbsH1hescPou5f1sc9081V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescPou5f1sc9081V0416102AlnRep2.bam	Link 1
RAD21	wgEncodeHaibTfbsH1hescRad21V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescRad21V0416102AlnRep2.bam	Link 1
RBBP5	wgEncodeBroadHistoneH1hescRbbp5a300109aStdAlnRep1.bam wgEncodeBroadHistoneH1hescRbbp5a300109aStdAlnRep2.bam	Link 2

(continued)

Table 1
(continued)

ChIP-seq experiment name	Download aligned reads file from ENCODE	
REST (NRSF)	wgEncodeHaibTfbsH1hescNrsfV0416102AlnRep1.bam wgEncodeHaibTfbsH1hescNrsfV0416102AlnRep2.bam	Link 1
RXRA	wgEncodeHaibTfbsH1hescRxraV0416102AlnRep1.bam wgEncodeHaibTfbsH1hescRxraV0416102AlnRep2.bam	Link 1
SIN3A	wgEncodeHaibTfbsH1hescSin3ak20Pcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescSin3ak20Pcr1xAlnRep2.bam	Link 1
SIX5	wgEncodeHaibTfbsH1hescSix5Pcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescSix5Pcr1xAlnRep2.bam	Link 1
SP1	wgEncodeHaibTfbsH1hescSp1Pcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescSp1Pcr1xAlnRep2.bam	Link 1
SP2	wgEncodeHaibTfbsH1hescSp2V0422111AlnRep1.bam wgEncodeHaibTfbsH1hescSp2V0422111AlnRep2.bam	Link 1
SP4	wgEncodeHaibTfbsH1hescSp4v20V0422111AlnRep1.bam wgEncodeHaibTfbsH1hescSp4v20V0422111AlnRep2.bam	Link 1
SRF	wgEncodeHaibTfbsH1hescSrfPcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescSrfPcr1xAlnRep2.bam	Link 1
TAF1	wgEncodeHaibTfbsH1hescTaf1V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescTaf1V0416102AlnRep2.bam	Link 1
TAF7	wgEncodeHaibTfbsH1hescTaf7sc101167V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescTaf7sc101167V0416102AlnRep2.bam	Link 1
TCF12	wgEncodeHaibTfbsH1hescTcf12Pcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescTcf12Pcr1xAlnRep2.bam	Link 1
USF1	wgEncodeHaibTfbsH1hescUsf1Pcr1xAlnRep1.bam wgEncodeHaibTfbsH1hescUsf1Pcr1xAlnRep2.bam	Link 1
YY1	wgEncodeHaibTfbsH1hescYy1sc281V0416102AlnRep1.bam wgEncodeHaibTfbsH1hescYy1sc281V0416102AlnRep2.bam	Link 1
H3K27ac	wgEncodeBroadHistoneH1hescH3k27acStdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k27acStdAlnRep2.bam	Link 2
H3K27me3	wgEncodeBroadHistoneH1hescH3k27me3StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k27me3StdAlnRep2.bam	Link 2
H3K36me3	wgEncodeBroadHistoneH1hescH3k36me3StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k36me3StdAlnRep2.bam	Link 2
H3K4me1	wgEncodeBroadHistoneH1hescH3k4me1StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k4me1StdAlnRep2.bam	Link 2
H3K4me2	wgEncodeBroadHistoneH1hescH3k4me2StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k4me2StdAlnRep2.bam	Link 2

(continued)

Table 1
(continued)

ChIP-seq experiment name	Download aligned reads file from ENCODE	
H3K4me3	wgEncodeBroadHistoneH1hescH3k4me3StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k4me3StdAlnRep2.bam	Link 2
H3K79me2	wgEncodeBroadHistoneH1hescH3k79me2StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k79me2StdAlnRep2.bam	Link 2
H3K9ac	wgEncodeBroadHistoneH1hescH3k9acStdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k9acStdAlnRep2.bam	Link 2
H3K9me3	wgEncodeBroadHistoneH1hescH3k09me3StdAlnRep1.bam wgEncodeBroadHistoneH1hescH3k09me3StdAlnRep2.bam	Link 2
H4K20me1	wgEncodeBroadHistoneH1hescH4k20me1StdAlnRep1.bam wgEncodeBroadHistoneH1hescH4k20me1StdAlnRep2.bam	Link 2

A collection of 31 TF and 10 HM ChIP-seq datasets (H1 Esc) from the ENCODE production phase is used for the analysis. The hg19 aligned bam files are downloaded from the following ENCODE Data Coordination Center links
Link 1: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/
Link 2: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/
Link 3: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromChip/

wgEncodeHaibTfbsH1hescAtf2sc81188V0422111AlnRep1.bam and wgEncodeHaibTfbsH1hescAtf2sc81188V0422111AlnRep2.bam.

3. At the end of this step, you must have 41 directories (31 TFs and 10 HMs) with two bam files each (with the exception of CTCF, where we have four bam files—*see* Table 1).
4. Use the ChIPseeqerSplitReadFiles script of the ChIPseeqer software [15] for each of the ChIP-seq datasets/directories independently to split both bam files into one reads file per chromosome (http://icb.med.cornell.edu/wiki/index.php/Elementolab/Split_raw_data). This step not only facilitates the analysis by transforming the sequence alignment reads into a platform independent data structure representing the ChIP-seq experiment but also combines multiple replicates of each experiment.

3.2 Preparation of RNA-seq Dataset

1. Unzip the gtf.gz file.
2. Extract the columns "Gene id," "RPKM1," and "RPKM2," which correspond to the normalized RNA-seq expression values from replicate 1 and replicate 2, respectively. Estimate the average RPKM for each gene.
3. At the end of this step, you should have a text file with two columns, separated by the TAB delimiter: the first column is the gene id, and the second column the average RPKM value from the two replicates (*see* **Note 1**).

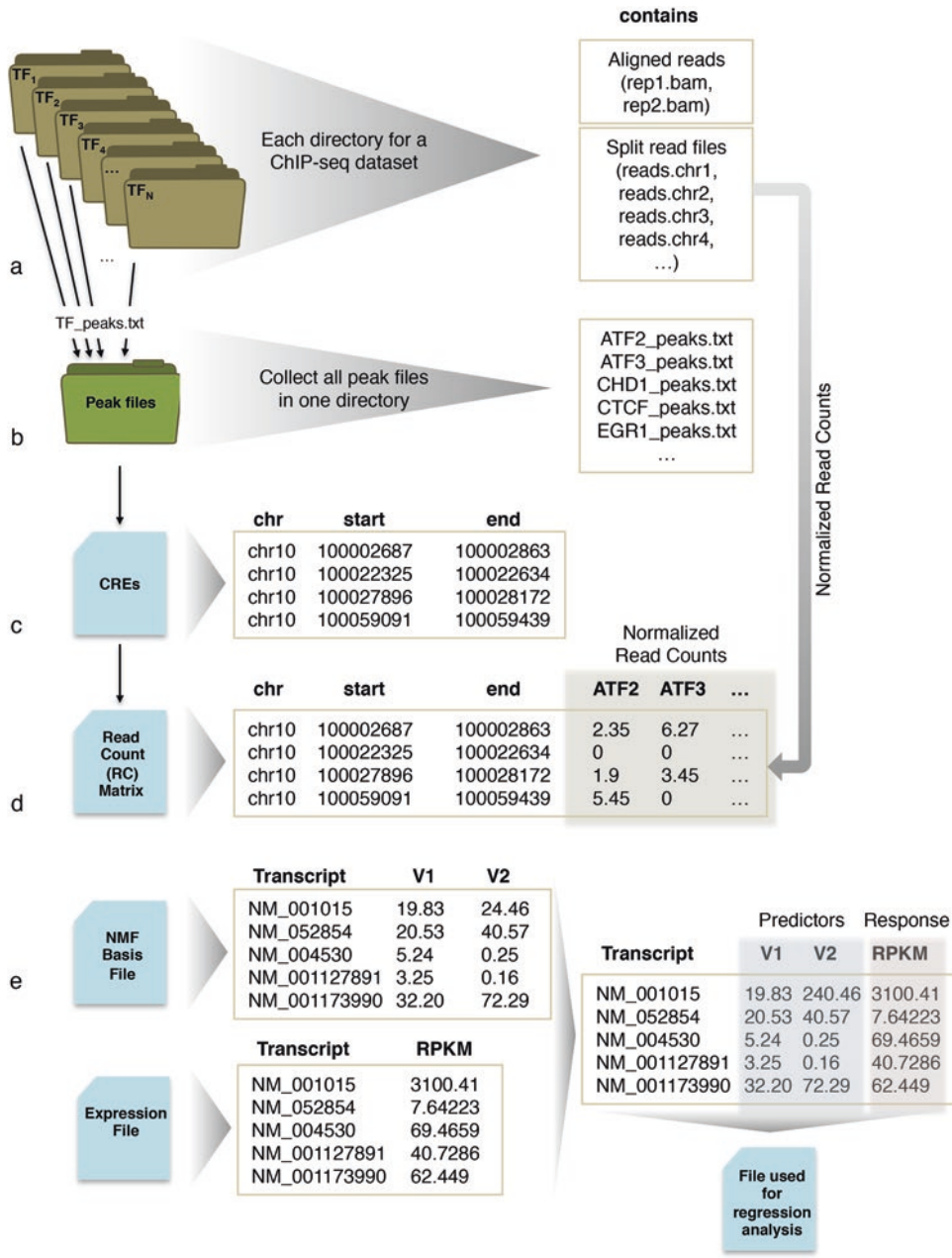

Fig. 2 Data file flow and description. (**a**) Collect aligned reads (bam files) for each ChIP-seq dataset. (**b**) Perform peak detection and collect all peak files in one directory. (**c**) Integrate all peaks into CREs. (**d**) Expand the CRE file by adding the normalized read counts for each ChIP-seq dataset (RC matrix). (**e**) Integrate the NMF basis file with the expression file to perform regression analysis

3.3 ChIP-seq Peak Calling

The process of identifying enriched TF binding sites, or locations of HMs, is known as ChIP-seq peak detection or peak calling (Fig. 1a).

1. In the described methodology, peak detection for each TF/HM dataset obtained previously from ENCODE is performed using the ChIPseeqer.bin program of the ChIPseeqer package [15] (http://icb.med.cornell.edu/wiki/index.php/Elementolab/ChIPseeqer_use).
2. Run the program using the same parameters for all datasets (i.e., $t = 10^{-5}$, *mindist* = 100 bp), except for broad domain modifications (e.g., H3K36me3, H4K20me1), where one parameter needs to be adjusted (*mindist* = 1000 bp) in order to capture wide peaks not as sharp as TF peaks (*see* **Note 2**).
3. At the end of this step, you should have one peak file for each TF/HM dataset.
4. Save all peak files into a new directory to prepare for the next step (Fig. 2b).

3.4 Identification of Multi-binding CREs

At this step, we merge the detected peaks from all experiments into CREs: regions with enrichment in at least one ChIP-seq dataset (Fig. 1b).

1. Use the CompareIntervalsMergedMultiEncode script available in ChIPseeqer [15] (http://icb.med.cornell.edu/wiki/index.php/Elementolab/CompareIntervalsMergedMulti) to integrate the ChIP-seq peak files identified earlier into multi-binding CREs (*see* **Note 3**).
2. At the end of this step, you obtain a peak file containing one CRE per line and the CRE's chromosome and start and end positions as three columns separated by the TAB delimiter (Fig. 2c).

3.5 Building the Read Count Matrix

For each CRE, we quantify the normalized ChIP-seq reads density in every experiment in order to build a read count matrix (RC matrix), whose rows correspond to CREs and columns to ChIP-seq experiments (Fig. 2d).

1. Use the ChIPseeqerReadCountMatrix tool available in ChIPseeqer [15] (http://icb.med.cornell.edu/wiki/index.php/Elementolab/ChIPseeqerGetReadCountInPeaksMatrix). This program performs RPKM-style read count normalization, so that multiple experiments with different numbers of reads are comparable, and quantifies the normalized reads for the CREs.
2. At the end of this step, you obtain the read count (RC) matrix, representing for every CRE (N rows) the reads density profiles of different ChIP-seq experiments (M columns) (Fig. 2d).

3.6 Perform NMF on Read Counts

The objective of NMF is to explain the observed data using a limited number of components, which when combined approximate the original data as accurately as possible. In particular, NMF

decomposes the RC matrix (N × M) into a basis matrix (N × k) and a mixture coefficient matrix (k × M) (Fig. 1c). The basis matrix has size N × k (each of the k columns defines a predicted complex) and contains the coefficient of each CRE in each complex [14]. In the coefficient matrix, the M columns represent the complex-binding pattern of the corresponding experiment [14].

Here we provide the steps for performing NMF analysis on the RC matrix and indicative R commands to assist the readers who want to replicate the NMF run.

1. Perform NMF on the RC matrix to group the CREs into clusters. Use the different built-in algorithms of the NMF package (e.g., brunet, nsNMF, offset), as well as the different seeding methods (e.g., random, ica) [16]. Note that when using a random seeding method, multiple runs are generally required to achieve stability and avoid bad local minima (*see* **Note 4**). In the example shown below, NMF runs with the "brunet" algorithm and the "random" seed method.
```
H1.table <- read.table("H1_reads_matrix.
txt", header=TRUE, row.names=1)
H1.mat <- data.matrix(H1.table)
H1.res.2.brunet.random <- nmf(H1.mat,
method="brunet", seed="random", 2)
```
2. Save the coefficients matrix as a heatmap.
```
pdf(file="H1_reads_matrix.txt.
H1.res.2.brunet.random.heatmap.pdf")
coefmap(H1.res.2.brunet.random)
dev.off()
```
3. Save the basis matrix as a text file. The coefficients in the basis matrix are protein complex scores that characterize each CRE. We use these scores at the next step, to model the regulatory effect of a complex on each TSS (i.e., gene).
```
H1.res.2.brunet.random.basis <- basis(H1.res.2.brunet.
random)
write.table(H1.res.2.brunet.random.basis,"H1_reads_matrix.
txt.H1.res.2.brunet.random.basis.txt", row.names=T, sep =
" ")
```

A critical parameter in this workflow is the factorization rank *r* for NMF, which defines the number of clusters used to approximate the original matrix and, therefore, the number of predicted complexes. In the example above, NMF is performed with $r = 2$. Given the NMF algorithm and the target matrix, a common way to estimate whether a given rank decomposes the original matrix into meaningful clusters is to try different values (e.g., *r* ranging from 2 to 20), compute some quality measures of the results that have been previously proposed for this type of approach [14], and choose the best value according to the quality criteria (*see* **Note 5**).

3.7 Measure the Influence of a Complex on a Gene

To model the effect that each protein complex has on a gene, we use the method described in [13] (implemented in the CalcExtendedPeakScores tool of ChIPseeqer) (*see* **Note 6**).

1. Identify all CREs that are within 50 kb of each TSS. Each CRE has one score per complex (stored in the basis matrix output of NMF) that quantifies the presence of each complex in a CRE.
2. For each CRE, estimate a new weighted score per complex by taking into account the genomic distance of a CRE to the corresponding TSS: the impact of a CRE to a TSS decreases exponentially with its genomic distance to the TSS (Fig. 1d) [13].
3. Sum the weighted scores per complex for all CREs that overlap with the 50 kb window of a TSS. The result is the Binding Influence Score (BIS) between a complex and a gene. Formally, this is modeled as follows:

$$\mathrm{BIS}_{i,j} = \sum_k h_k e^{-d_k/d_0}$$

In this formula, BIS quantifies the interaction between gene j and protein complex i. h_k is the complex score of CRE k and d_k is the distance between the TSS and the CRE. d_0 is a constant used in the ratio d_k/d_0 to specify the shape of the exponential function [13]. The larger the d_0, the more distal complexes will influence the TSS and its BIS score. Here, we set d_0 to 5000 bp (*see* **Note 7**).

4. At the end of this step, you obtain a text file, representing for every transcript (rows) the BIS scores of each complex (Fig. 2e). The first six lines of the new file should look like this:

Gene	V1	V2
NM_001015	19.83	240.46
NM_052854	20.53	40.57
NM_004530	5.24	0.25
NM_001127891	3.25	0.16
NM_001173990	32.20	72.29

3.8 Integration of RNA-seq

1. Merge the file you created from the previous step, with the RNA-seq expression file that contains the RPKM values for each transcript (*see* **Note 8**).
2. At the end of this step you obtain a text file with transcript names as rows, the BIS scores of each complex as columns, and a new column that corresponds to the RPKM value of the transcript (Fig. 2e). The first six lines of the new file should look like this:

Gene	V1	V2	RPKM
NM_001015	19.83	240.46	3100.41
NM_052854	20.53	40.57	7.64223
NM_004530	5.24	0.25	69.4659
NM_001127891	3.25	0.16	40.7286
NM_001173990	32.20	72.29	62.449

3.9 Application of Regression Model

The BIS values (columns V1 and V2 in the file example above) are then used as explanatory variables, or predictors, to assess the contribution of a detected protein complex to gene expression or response (column RPKM in the file example above).

Here we provide the steps for performing linear regression analysis in R (function lm) and indicative R commands to assist the readers who want to replicate the steps.

1. Perform linear regression using the model:

$$\mathrm{mRNA}_j = \beta_0 + \sum_i^m \beta_i \mathrm{BIS}_{ij} + \varepsilon_j$$

In this formula, mRNA_j is the absolute mRNA expression value of gene j, and BIS_{ij} is the score of gene j in complex i. The $\hat{\beta}_i$ coefficients are estimated using ordinary least square fitting, and their statistical significance is determined using the t-test.

```
m <- read.csv("H1_reads_matrix_broadHM.txt.
H1.res.2.brunet.random.basis.txt. H1_RPKM.
txt",header=T, row.names=1, sep="\t", check.
names=T)
attach(data.frame(m))
fit <- lm(log(RPKM+1) ~ V1 + V2)
summary(fit)
```

Importantly, using the log-transformed response variable (RPKM) makes it easier to interpret the exponentiated regression coefficients.

2. The R output at the end of this step contains the $\hat{\beta}_i$ coefficients of the complexes, as well as the coefficient of determination (R^2). A significant and positive $\hat{\beta}_i$ coefficient indicates that the corresponding protein complex contributes positively to the mRNA expression values. On the other hand, a negative coefficient indicates negative (i.e., repressive) contribution. The coefficient of determination R^2 measures the quality of the overall fit of the model and indicates the proportion of the gene expression variation explained by the model.

3. Evaluate the performance of the regression model by estimating the prediction accuracy using Spearman correlation between actual and predicted gene expression values.

   cor.test(predict(fit, m), log(RPKM+1), method="spearman")

4 Notes

1. Although there are numerous ways to extract the required columns from a text file in a Unix-based OS, we provide the following command to assist the readers who want to replicate the method:

   ```
   cat GSM758566_hg19_wgEncodeCshlLongRnaSeqH-
   1hescCellPapGeneGencV7.gtf | cut -f9 | cut
   -d ";" -f 1,3,4 | cut -d "\"" -f 2,4,6 | sed
   's/\"/<TAB>/g' | awk -F"\t" '{print $1"\t"
   ($2+$3)/2}' > GSM758566_hg19_H1Esc_RNAseq.txt
   ```

   This command outputs the "Gene id" columns, as well as the average of columns "RPKM1" and "RPKM2," into a new file named "GSM758566_hg19_H1Esc_RNAseq.txt." The first six lines of the new file should look like this:

ENSG00000174177.7	6.01899
ENSG00000225538.1	0
ENSG00000237851.1	0
ENSG00000243765.1	0
ENSG00000203388.2	0.0109565
ENSG00000151503.7	9.16775

   In order to be consistent with the RefSeq annotation used in the pipeline (*see* **Note 6**), we suggest converting the Gencode IDs to RefSeq NMs: first trim the ending of the Gencode IDs to get Ensembl IDs (e.g., ENSG00000174177.7 → ENSG00000174177), and then use the BiomaRt package in R to convert the Ensembl IDs to the corresponding RefSeq ones. Indicatively, provide the following R commands to perform the Ensembl to RefSeq annotation conversion:

   ```
   RPKM<- read.table(file="GSM758566_hg19_wgEnco-
   deCshlLongRnaSeqH1hescCellPapGeneGencV7
   .gtf", header=F)
   ensembl<- useMart("ensembl",
   dataset="hsapiens_gene_ensembl")
   values <- RPKM_E$V10
   getBM(attributes=c("refseq_mrna", "ensembl_
   gene_id", "hgnc_symbol"), filters = "ensembl_
   gene_id", values = values, mart= ensembl)
   ```

2. In order to capture peaks of broad domain histone modifications (e.g., H3K36me3, H3K79me2), we suggest tuning the following ChIPseeqer.bin parameters:

   (a) *t*: significance negative log *p*-value threshold for peaks. For example, $t = 15$ indicates 10^{-15}.

   (b) *mindist*: minimum distance between peaks (merge subpeaks otherwise).

   Using lower *t* threshold (such as $t = 5$) allows including peaks that are not very "sharp", while increasing *mindist* value (such as 1000, 10,000) allows merging continuous enriched regions into a large peak.

3. Alternatively, the identification of CREs can be performed by using the intersect or intersectBed tools from the BEDTools suite [17].

4. The stochastic nature of the seeding method used to compute the starting point of the chosen algorithm requires multiple NMF runs to achieve stability. The NMF R package gives the option to perform multiple runs with random initializations for the basis and coefficient matrices and keep the factorization that achieves the lowest approximation error across the multiple runs [16] (option nrun). We recommend the default 30 runs for a faster NMF analysis and 100 runs for better results.

5. The selection of rank r is of great importance when applying NMF. We recommend using the function nmfEstimateRank from the NMF package to estimate the quality measures for each rank *r*. nmfEstimateRank performs multiple NMF runs for a range of rank of factorization and, for each, returns a set of quality measures together with the associated consensus matrix. For example, the dispersion and the cophenetic correlation coefficients are both based on the consensus matrix (i.e., the average of connectivity matrices) and measure the stability and reproducibility of the clusters obtained from NMF for a certain value of *k*, respectively [14]. The explained variance measure evaluates how well the NMF model reconstructs the original data, while the sparseness measure, for both the basis and mixture coefficient matrices, shows whether an NMF representation encodes much of the data using only few components [5, 18]. We highly recommend using nmfEstimateRank to identify local maxima in these coefficients at high rank factorization, in order to find complexes with high granularity (*see* Supplementary Material in [5]).

6. The steps described in Subheadings 3.6, 3.7, 3.8 and 3.9 are included in the ChIPseeqer script NMFCall_Split.pl. It requires an RC matrix (Fig. 2d) and a transcript-based expression file (Fig. 2e). We provide the following command to assist the readers who want to replicate the method:

```
NMFCall_Split.pl --matrix=H1_reads_matrix_
broadHM.txt --expfile=H1_RPKM.txt
--response=log"(RPKM+1)" --label=H1
```

Note that the RefSeq annotation is used in this pipeline.

7. The d_0 parameter choice leads to a rapidly decreasing exponential function, which strongly penalizes distal regulatory elements (*see* Supplementary Material in [5]). We recommend increasing the value if interested to give higher BIS scores to complexes that we expect to bind to CREs far away from the TSS (e.g., the "enhanceosome" complex).
8. This step is included in the NMFCall_Split.pl script (*see* **Note 6**). Alternatively, to merge the two files on a specified field, you can use the online tool "Join two Datasets" available in Galaxy (https://usegalaxy.org/).

Acknowledgments

E.G. is supported by start-up finds provided by the City University of New York and the Hospital for Special Surgery. O.E. is supported by NSF CAREER, LLS SCOR, Hirschl Trust Award, Starr Cancer Consortium I6-A618, NIH 1R01CA194547.

References

1. Wittkopp PJ, Kalay G (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13(1):59–69. doi:10.1038/nrg3095, nrg3095 [pii]
2. Siepel A, Arbiza L (2014) Cis-regulatory elements and human evolution. Curr Opin Genet Dev 29:81–89. doi:10.1016/j.gde.2014.08.011, S0959-437X(14)00092-6 [pii]
3. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng J, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M (2012) Architecture of the human regulatory network derived from ENCODE data. Nature 489(7414):91–100. doi:10.1038/nature11245, nature11245 [pii]
4. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473(7345):43–49. doi:10.1038/nature09906, nature09906 [pii]
5. Giannopoulou EG, Elemento O (2013) Inferring chromatin-bound protein complexes from genome-wide binding assays. Genome Res 23(8):1295–1306. doi:10.1101/gr.149419.112, gr.149419.112 [pii]
6. Ji H, Li X, Wang QF, Ning Y (2013) Differential principal component analysis of ChIP-seq. Proc Natl Acad Sci U S A 110(17):6789–6794. doi:10.1073/pnas.1204398110, 1204398110 [pii]
7. Jiang P, Singh M (2014) CCAT: Combinatorial Code Analysis Tool for transcriptional regulation. Nucleic Acids Res 42(5):2833–2847. doi:10.1093/nar/gkt1302, gkt1302 [pii]
8. Ram O, Goren A, Amit I, Shoresh N, Yosef N, Ernst J, Kellis M, Gymrek M, Issner R, Coyne M, Durham T, Zhang X, Donaghey J, Epstein CB, Regev A, Bernstein BE (2011) Combinatorial patterning of chromatin regu-

lators uncovered by genome-wide location analysis in human cells. Cell 147(7):1628–1639. doi:10.1016/j.cell.2011.09.057, S0092-8674(11)01448-6 [pii]

9. Whitington T, Frith MC, Johnson J, Bailey TL (2011) Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Res 39(15), e98. doi:10.1093/nar/gkr341, gkr341 [pii]
10. Wong KC, Li Y, Peng C, Zhang Z (2015) SignalSpider: probabilistic pattern discovery on multiple normalized ChIP-Seq signal profiles. Bioinformatics 31(1):17–24. doi:10.1093/bioinformatics/btu604, btu604 [pii]
11. Xie D, Boyle AP, Wu L, Zhai J, Kawli T, Snyder M (2013) Dynamic trans-acting factor colocalization in human cells. Cell 155(3):713–724. doi:10.1016/j.cell.2013.09.043, S0092-8674(13)01217-8 [pii]
12. Zeng X, Sanalkumar R, Bresnick EH, Li H, Chang Q, Keles S (2013) jMOSAiCS: joint analysis of multiple ChIP-seq datasets. Genome Biol 14(4):R38. doi:10.1186/gb-2013-14-4-r38, gb-2013-14-4-r38 [pii]
13. Ouyang Z, Zhou Q, Wong WH (2009) ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. Proc Natl Acad Sci U S A 106(51):21521–21526. doi:10.1073/pnas.0904863106, 0904863106 [pii]
14. Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 101(12):4164–4169. doi:10.1073/pnas.0308531101
15. Giannopoulou EG, Elemento O (2011) An integrated ChIP-seq analysis platform with customizable workflows. BMC Bioinformatics 12:277. doi:10.1186/1471-2105-12-277
16. Gaujoux R, Seoighe C (2010) A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11:367. doi:10.1186/1471-2105-11-367, 1471-2105-11-367 [pii]
17. Quinlan AR (2014) BEDTools: The Swiss-Army tool for genome feature analysis. Curr Protoc Bioinformatics 47:11.12.11–11.12.34. doi:10.1002/0471250953.bi1112s47
18. Pascual-Montano A, Carazo JM, Kochi K, Lehmann D, Pascual-Marqui RD (2006) Nonsmooth nonnegative matrix factorization (nsNMF). IEEE Trans Pattern Anal Mach Intell 28(3):403–415. doi:10.1109/TPAMI.2006.60

Chapter 5

Assay for Transposase-Accessible Chromatin with High-Throughput Sequencing (ATAC-Seq) Protocol for Zebrafish Embryos

Canan Doganli, Melissa Sandoval, Sean Thomas, and Daniel Hart

Abstract

Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) is a useful method to map genome-wide chromatin accessibility and nucleosome positioning. Genome-wide sequencing is performed utilizing adapter sequences inserted by a prokaryotic transposase, Tn5, into the accessible regions of chromatin. Here we describe the use of ATAC-seq in the zebrafish embryo and thereby the applicability of this approach in whole vertebrate embryos.

Key words ATAC-seq, Chromatin accessibility, Genome wide, Zebrafish

1 Introduction

Eukaryotic cells package DNA by wrapping it around histones to make nucleosomes connected by DNA linker regions, which are further condensed to form chromatin [1]. Histones are subject to various posttranslational modifications that contribute to different chromatin states (i.e., euchromatin or heterochromatin), which in turn influence gene expression [2, 3]. Different chromatin states are correlated with differential gene expression. Heterochromatin, described as closed or inaccessible, is associated with negative regulation of gene expression. By contrast, euchromatin (open or accessible) is associated with positive gene expression.

Chromatin accessibility has been previously studied by probing DNA with enzymes (DNase I hypersensitivity mapping), mechanically shearing after cross-linking (chromatin immunoprecipitation, ChIP), and more recently by genome-wide integration of transposase. Genome-wide mapping of the regulatory landscape will significantly contribute to the understanding of transcriptional gene regulation.

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_5,

Assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) is a method developed by Greenleaf and colleagues used to map genome-wide chromatin accessibility and nucleosome positioning [4]. They adopted the use of a hyperactive prokaryotic transposase, Tn5, which has been previously mutated to significantly increase transposase [5]. Tn5 inserts a 19 base pair paired-end sequence (5′-CTGACTCTTATACACAAGT-3′) into accessible DNA. By taking advantage of Tn5 transposition into open regions of the genome together with the use of barcode sequences inserted into these regions, we can use the accessibility of DNA and nucleosome positioning as a new way to perform genome-wide chromatin accessibility assays.

The use of ATAC-seq has advantages over other methods because of the significantly reduced number of cells required as input, time savings of the assay, high-resolution mapping of chromatin accessibility, DNA footprinting, and nucleosome positioning. The use of zebrafish as a model organism coupled with ATAC-seq is a powerful combination to study chromatin dynamics in the context of the developing vertebrate embryo. Here we detail the ATAC-seq protocol performed on zebrafish embryos adapted from the report of Greenleaf and colleagues [6].

2 Materials

All solutions should be prepared using autoclaved ultrapure water (resistivity levels of 18.2 MΩ.cm at 25 °C).

2.1 Embryo Dissociation and Cell Preparation Components

1. Tweezers.
2. 100-mm petri dish.
3. Incubator (28.5 °C).
4. 1.5-ml microfuges.
5. 15-ml falcon tubes.
6. Micro pestle.
7. 1× E3 embryonic medium: 5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl, 0.33 mM $MgSO_4$ in H_2O. Make 60× E3 buffer by dissolving 34.8 g NaCl, 1.6 g KCl, 5.8 g $CaCl_2{\cdot}2H_2O$ and 9.78 g $MgCl_2{\cdot}6H_2O$ in 2 L of H_2O. Make 1:60 dilution of 60× stock by mixing 16.7 ml 60× E3 and 983.3 ml H_2O. Add 100 μl of 1% methylene blue as a fungicide to 1 L of medium.
8. 1× phosphate-buffered saline (PBS).
9. Lysis buffer: 10 mM Tris–HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40. Prepare 1 M stock solution of Tris–HCl by dissolving 121.1 g of Tris base in 800 ml H_2O. Adjust pH to 7.4 by adding concentrated HCl. Adjust the volume of the

solution to 1 L with H_2O. Make 1 M NaCl and 1 M $MgCl_2$ by dissolving 5.8 g NaCl and 20.33 g $MgCl_2$, respectively, in 100 ml H_2O. Make a 1:10 dilution of NP-40 by mixing 1 ml of NP-40 and 9 ml H_2O (10% NP-40). Combine 100 µl of 1 M Tris–HCl, pH 7.4, 100 µl of 1 M NaCl, 30 µl of 1 M $MgCl_2$, and 100 µl of 10% NP-40 in a 15 ml falcon tube, and fill with H_2O up to a total volume of 10 ml.

2.2 Transposition and PCR Amplification Components

1. Heating block.
2. Thermocycler.
3. Microcentrifuge.
4. 0.2-ml PCR tubes.
5. MicroAmp fast 96-well reaction plate (0.1 ml).
6. MicroAmp optical adhesive.
7. Real-Time PCR System.
8. Illumina Nextera DNA Sample Prep Kit (FC-121-1030).
9. Qiagen MinElute Reaction Cleanup Kit (28204).
10. SYBR Green I nucleic acid gel stain (Life Technologies, S7563): Prepare 100× SYBR Green I solution by mixing 1 µl of 10,000× SYBR Green I with 99 µl DMSO.
11. NEBNext High-Fidelity 2× PCR Master Mix (New England Labs, M0541).

3 Methods

3.1 Embryo Dissociation and Cell Preparation

1. Collect zebrafish embryos in a 100-mm petri dish, and raise in 1× E3 embryo medium in an incubator set to 28.5 °C until desired stage.
2. Take a number of embryos with an approximate total cell number of 50,000 (*see* **Note 1**). For example, 50 embryos at 3 hpf (1000-cell stage), remove the chorions manually using tweezers (*see* **Note 2**). Deyolking is not necessary in our experience.
3. Collect dechorionated embryos in a 1.5-ml microcentrifuge tube and dissociate using a micro pestle and follow by mixing using a 200-µl pipette (*see* **Note 3**). Wash the pestle into the tube with 1× E3 embryo medium to collect all the cells on the pestle.
4. Centrifuge immediately at 500 × *g*, 4 °C for 5–10 min.
5. Remove supernatant without disturbing the cell pellet. Add 50 µl of cold 1× PBS and centrifuge at 500 × *g*, 4 °C for 5 min.
6. Remove supernatant and add 50 µl of cold lysis buffer. Gently flick the tube and then centrifuge immediately at 500 × *g*, 4 °C for 10 min. Place the tube on ice.

3.2 Transposition

1. Prepare 50 µl of transposition mix per reaction combining 25 µl Tagment DNA buffer, 2.5 µl Tagment DNA enzyme 1, and 22.5 µl nuclease-free H_2O. Add this to the lysed cells and gently mix by pipetting.
2. Incubate at 37 °C for 30 min on a heating block.
3. Clean reaction using Qiagen MinElute Reaction Cleanup Kit. Perform centrifugations at room temperature and at 12,000 × *g*.
4. Add 300 µl Buffer ERC to the reaction and mix. Place a MinElute column in a 2-ml collection tube, load the sample to the column, and centrifuge for 1 min. Empty the collection tube.
5. Add 750 µl Buffer PE to the MinElute column and centrifuge for 1 min. Empty the collection tube.
6. Centrifuge for 2 min to remove residual ethanol from the PE buffer.
7. Place the MinElute column in a clean 1.5-ml microcentrifuge tube. Add 10 µl of elution buffer to the column, incubate for 1 min at room temperature, and centrifuge for 1 min to elute DNA. Purified DNA sample can be stored at –20 °C at this point.

3.3 PCR Amplification

1. For PCR amplification, mix the reagents below in a PCR tube:

   10 µl of nuclease-free H_2O.

   10 µl of transposed DNA.

   2.5 µl of Nextera PCR Primer 1.

   2.5 µl of Nextera PCR Primer 2 (barcode).

   25 µl of NEBNext High-Fidelity 2× PCR Master Mix.

   *For sequences of the primers, please *see* Goryshin et al., 1998 [5].

   Run the PCR reaction:

   72 °C, 5 min

   98 °C, 30 s
   5 cycles:

   98 °C, 10 s.

   63 °C, 30 s.

   72 °C, 1 min.
   Hold at 4 °C.
2. Determine cycle number using qPCR. Prepare 15 µl of qPCR master mix:

   5 µl of 5 cycle PCR-amplified DNA.

   3.9 µl of nuclease-free H2O.

   0.25 µl of Nextera PCR Primer 1.

   0.25 µl of Nextera PCR Primer 2.

0.6 μl of 100× SYBR Green I.

5 μl of NEBNext High-Fidelity 2× PCR Master Mix.

3. Add reaction mix to a 96-well reaction plate and seal with optical adhesive film.
4. Run the qPCR reaction:

   98 °C, 30 s.

   20 cycles.

   98 °C, 10 s.

   63 °C, 30 s.

   72 °C, 1 min.

   Hold at 4 °C.
5. Determine the number of cycles to run for the remaining PCR reaction. For this, plot linear fluorescence vs. cycle in the StepOne software (Applied Biosystems), and make note of the cell-cycle number at which SYBR fluorescence intensity is half of the maximum fluorescence (*see* Fig. 1).
6. Return the rest of the PCR reaction back to the qPCR machine and run PCR reaction with the determined cycle number.

   98 °C, 30 s.

   Use number of cycles as determined in step 6.

   98 °C, 10 s.

   63 °C, 30 s.

   72 °C, 1 min.
7. Clean reaction using Qiagen MinElute Reaction Cleanup Kit as described in **steps 3–6** in Subheading 3.2.

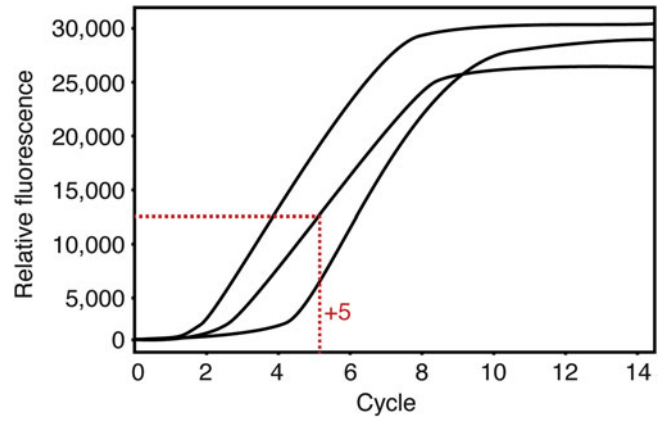

Fig. 1 A schematic RT-PCR amplification plot illustrating how to determine the number of cycles to add to the PCR amplification of the ATAC-seq libraries (**step 6**)

8. Place the MinElute column in a clean 1.5-ml microcentrifuge tube. Add 25 μl of elution buffer to the column, incubate for 5 min at room temperature, and centrifuge for 1 min to elute DNA. Purified DNA sample can be stored at −20 °C.
9. You can check sample quality using Bioanalyzer DNA 1000.

3.4 Sequencing and Analysis Considerations

1. *Sequencing depth.* Low-depth sequencing consistently reveals the most accessible genomic regions but leaves weakly accessible regions undiscovered. However, it is currently impractical to generate enough sequence depth to fully saturate the entire dynamic range of genome-wide accessibility. Therefore, great care should be put into determining the depth of sequencing to be used for each replicate based on the goals of the particular experiment.
2. *Replicates.* Often the most difficult choice to make concerns the balance between number of biological replicates and sequencing depth. Many of the most accessible sites will be found in any cell line consistently. It is especially important for lower-accessible sites to have both enhanced coverage and enough replicates to assess biological variability.
3. *Paired-end vs. single-end reads.* While paired-end sequencing is not necessary for establishing the locations of accessible chromatin and single-end sequencing will yield more coverage for the same cost, paired-end sequencing is helpful for identifying regions with more precisely positioned nucleosomes. The choice of whether to use single- or paired-end sequencing should be made accordingly.
4. *Read length.* 50 bp reads map uniquely to the vast majority of the genome. Given the limitations of most sequencing budgets, it often makes more sense to increase depth of coverage than to increase sequence length, as mappability does not increase dramatically as reads get longer than 50 bp.
5. *Quality control.* Once reads have been mapped to the genome of choice, for most projects, only fragments that map uniquely to the genome (mapq ≥ 30) should be used. To track quality, it is often useful to take note of the percent of tags that map uniquely to the genome in addition to the fraction of reads that represent mitochondrial "contamination."
6. *Integration location.* The exact integration location represented by each read can be identified as the 5′-most position of reads that map to the reference strand +4 bp. For reads that map to the non-reference strand, the location is the 3′-most position of the read −5 bp.
   For an example of successful ATAC-seq in zebrafish embryos, *see* Fig. 2.

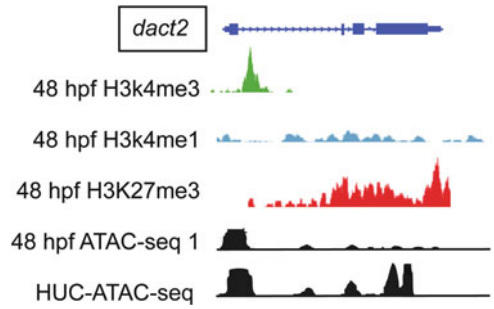

Fig. 2 ATAC-seq in whole zebrafish embryos and isolated neurons. Strong agreement between ATAC-seq peaks and histone modification data from Gomez-Skarmeta lab. The data reveal distinct ATAC peaks in the HuC neurons sorted from *Tg*(*HuC:GFP*) zebrafish line, compared to rest of the embryo, at 48 hpf

4 Notes

1. When working with embryonic stages with unknown cell number, trypan blue staining can be used to get an estimate of the total viable cell number. Place a known number of embryos in a 1.5-ml microcentrifuge tube. Remove embryo medium and add 1 ml deyolking buffer (55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$). Pipette up and down to dissolve yolk. Centrifuge at 500 × *g* for 2 min, and discard supernatant. Wash the pellet in 200 µl of 1× E3 embryo medium, centrifuge at 500 × *g* for 2 min, and discard the supernatant. Resuspend pellet in 200 µl of 1× E3 embryo medium and dissociate cells using micro pestle and pipetting. If not fully dissociated, perform trypsin dissociation as following. Centrifuge cells at 500 × *g* for 2 min and discard supernatant. Resuspend the pellet in trypsin–EDTA solution (0.5 mg/ml trypsin and 0.22 mg/ml EDTA, Ca^{2+}- and Mg^{2+}-free) and incubate for 5–7 min. Stop the dissociation by adding Hi-FBS to a final concentration of 5%. Microcentrifuge at 500 × *g* for 5 min. Discard the supernatant and add 200 µl of FACSMax cell dissociation solution (Genlantis, T200100) or 1× PBS. Mix 5 µl of the cell mix with 5 µl of 4% trypan blue solution. Incubate at room temperature for 5 min, load mix to a Countess Cell Counting Chamber Slide (Invitrogen), and get cell count using Countess automated cell counter (Invitrogen). Counting cells can also be done manually using hemocytometer. Perform cell counts three times and use the average number to get a total cell number estimate.
2. If using a large number of embryos, dechorionating can be performed enzymatically using pronase treatment. Prepare 1% (w/v) pronase (protease from *Streptomyces griseus*) solution:

Dissolve 1 g of pronase in 100 ml of 1× E3 embryo medium, incubate for 2 h at 37 °C, aliquot, and store at −20 °C. To remove chorions, place embryos in a 100-ml beaker. Add pronase (1 % w/v) to a dilution of 1:10 depending on the medium volume in the beaker. Incubate for 5–10 min at 28.5 °C swirling occasionally. Incubation time will vary depending on the developmental stage of the embryos, shorter times for young embryos, determined empirically. When chorions are mostly removed, fill the beaker with 1× E3 embryo medium. Discard medium and wash with 1× E3 several times until all traces of pronase have been removed.

3. ATAC-seq can be performed similarly on sorted cells. Cell dissociation for sorting can be performed by trypsin digestion as in **Note 1**. Cell sorting steps are adapted from Manoli and Driever (2012) [6]. Briefly, after stopping trypsin dissociation by adding Hi-FBS to a final concentration of 5 %, microcentrifuge at 500×*g* for 5 min. Discard the supernatant and add 500 µl of FACSMax cell dissociation solution. Moisten a cell strainer (with 40-µm mesh) with FACSMax solution, and place it on a 100-mm petri dish on ice. Add the cell and mix into the cell strainer. Use the plunger from a 1-ml syringe and carefully pass the cells through the strainer by pressing the plunger over the cells. Collect the cell suspension from petri dish into a 5-ml round bottom polystyrene test tube (Corning, 352058). Try to get all the cells from the petri dish by adding extra FACSMax cell dissociation solution. Place the cell suspensions on ice and take to the flow cytometry facility. Upon sorting perform ATAC-seq from lysis step (*see* Subheading 3.1, **steps 4–6**) and onwards.

References

1. Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. Science 184(4139):868–871
2. Kouzarides T (2007) Chromatin modifications and their function. Cell 128(4):693–705. doi:10.1016/j.cell.2007.02.005, S0092-8674(07)00184-5 [pii]
3. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128(4):707–719. doi:10.1016/j.cell.2007.01.015, S0092-8674(07)00109-2 [pii]
4. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10(12):1213–1218. doi:10.1038/nmeth.2688, nmeth.2688 [pii]
5. Goryshin IY, Reznikoff WS (1998) Tn5 in vitro transposition. J Biol Chem 273(13):7367–7374
6. Manoli M, Driever W (2012) Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. Cold Spring Harb Protoc 2012(8). doi:10.1101/pdb.prot069633, 2012/8/pdb.prot069633 [pii]

Chapter 6

Establishment of Time- and Cell-Specific RNAi in *Caenorhabditis elegans*

Masayuki Hamakawa and Takaaki Hirotsu

Abstract

The nematode worm *Caenorhabditis elegans*, in which loss-of-function mutants and RNA interference (RNAi) models are available, is a model organism useful for analyzing effects of genes on various life phenomena. In particular, RNAi is a powerful tool that enables time- or cell-specific knockdown via heat shock-inducible RNAi or cell-specific RNAi. However, the conventional RNAi methods are insufficient for investigating pleiotropic genes with various sites of action and life stage-dependent functions. To investigate the temporal- and cell-specific profiles of multifunctional genes, we established a new RNAi method that enables simultaneous time- and cell-specific knockdown (T.C.RNAi) in *C. elegans*. In this method, one RNA strand is expressed by a cell-specific promoter and the other by a heat shock promoter, resulting in only expression of double-stranded RNA in the target cell when heat shock is induced. We confirmed the effect of T.C.RNAi by the knockdown of GFP and the *odr-3* gene which encodes Gα and is essential for olfaction. Further, this technique revealed that the control of glutamate receptors GLR-1 localization in RMD motor neurons requires Ras at the adult stage to regulate locomotion behavior.

Key words RNAi, Conditional knockdown, Time and cell specific, Heat shock, *C. elegans*

1 Introduction

The nematode *Caenorhabditis elegans* is useful for studying the functions of genes through facilitated analyses. In this organism, various cell-specific promoters can be utilized for cell-specific expression to determine the function of a gene in a specific cell. Recently, temporal control of cell-specific expression using heat shock factor-1 (*hsf-1*) mutants has been reported [1]. In addition, the effects of gene knockdown can be assessed using various loss-of-function mutants, and researchers recently developed a new method of generating loss-of-function mutants in targeted genes in *C. elegans* [2].

RNAi is one of the most powerful tools for gene knockdown. RNAi-mediated cell-specific knockdown in *C. elegans* is a currently available technique based on driving the expression of double-stranded

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_6,

RNA in target cells via cell-specific promoters [3]. This method identifies the cells in which the target gene functions. Moreover, gene functions can be analyzed by this system because of cell specificity, even if mutants of the gene show fatal phenotypes. In contrast to cell-specific promoters, heat shock promoters drive the expression of genes at arbitrary timing via heat shock [4]. A previous report has shown that RNA hairpins, which are driven by the heat shock promoter, induce knockdown of the target gene [5], suggesting the possibility of time-specific knockdown. However, heat shock promoters drive global expression, meaning that the knockdown is performed in the majority of cells, thereby removing vital cellular functions and thus inducing lethality when essential genes are knocked down by this method. Therefore, for detailed analysis of multifunctional genes, such as the components of the Ras-MAPK pathway, simultaneous time-specific and cell-specific knockdown is necessary. To our knowledge, however, such a method has not yet been developed in *C. elegans*.

We developed a novel RNAi method which can provide simultaneous time- and cell-specific knockdown. The use of two different types of promoters, the heat shock promoter and the cell-specific promoter, enables the time- and cell-specific expression of double-stranded RNA. We named the novel method time- and cell-specific RNAi (T.C.RNAi). The effect of gene knockdown by T.C.RNAi was confirmed via knockdown of GFP. In addition, we actually revealed when and where Ras functions for the control of locomotion behavior using T.C.RNAi, indicating T.C.RNAi is a useful method providing gene knockdown at arbitrary timing only in a target cell.

2 Materials

2.1 Worm Culture

C. elegans strains were cultured on NGM plates (at 20 °C, under standard conditions) with *Escherichia coli* NA22 as the food source [6].

2.2 Germ-Line Transformation

Transgenic lines were generated by microinjection as previously described [7]. *lin-44p*::GFP and *myo-3p*::GFP were used as transformation markers.

2.3 Chemotaxis Assay

9 cm plates with the assay format described in Fig. 3a were used. The composition of assay plates was 20 g/L Bacto agar, 5 mM KPO_4, 1 mM $CaCl_2$, and 1 mM $MgSO_4$.

2.4 Heat Shock Treatment

A heat block was used for heat shock treatment for shorter than 30 min, such as heat shock at the adult stage. An incubator was used for heat shock over 1 h, such as heat shock at the embryonic or larval stage.

2.5 Heat Shock Promoter and Cell-Specific Promoters

The sequence data and primers of each promoter used in this study are referred from the references described below. The *myo-3* promoter and the heat shock promoter *hsp16-2* were referred from Fire

Lab. vectors. Information of other cell-specific promoters was obtained from WormBase (http://www.wormbase.org/). These promoters were amplified with forward and reverse primers. At the 5′ end of the reverse primer, 25 nt sequence complementary to the 3′ or 5′ end of the target region was attached, and the promoter region and the target region were combined by PCR as described in [3].

3 Methods

3.1 The Mechanism of Time- and Cell-Specific RNAi

To enable simultaneous time-specific and cell-specific knockdown, we modified the previous method, cell-specific RNAi, in which cell-specific knockdown is achieved by the expression of double-stranded RNA by the cell-specific promoter [3]. The heat shock promoter *hsp16-2* [4], which drives the expression in almost all tissues, is able to activate gene expression at arbitrary timing by heat shock (*see* **Note 1**). By introducing the heat shock promoter to the cell-specific RNAi technique, we established a new method for time-specific and cell-specific knockdown of genes, in which the expression of one RNA strand (sense or antisense) was driven by a cell-specific promoter and the expression of the other RNA strand was induced by a heat shock promoter (Fig. 1). Under the normal condition (20–24 °C), single-stranded RNA was expressed in the target cells by the cell-specific promoter, whereas under the heat shock-inducing condition (30–33 °C), double-stranded RNA was expressed in only target cells by the heat shock promoter and the cell-specific promoter. This method was termed time-specific and cell-specific RNAi (T.C.RNAi).

3.2 Protocol of T.C.RNAi

This paragraph shows the basic protocol of T.C.RNAi. Detailed conditions such as the expression level of sense and antisense RNA strands and the quantification of gene knockdown are different according to a target gene. As a reference, you should read the paragraphs of 3.3. Knockdown of GFP and 3.4. Knockdown of endogenous genes

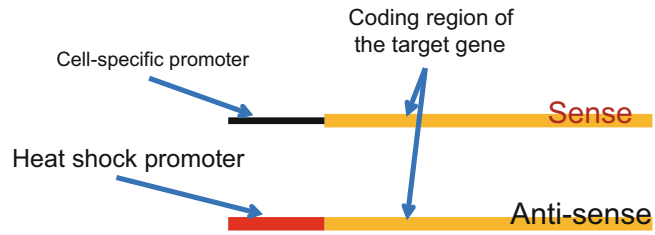

Fig. 1 Outline of time- and cell-specific RNAi. Under the standard condition (20 °C), expression of only single-stranded RNA is driven by the cell-specific promoter, whereas under the heat-shocked condition (30–33 °C), double-stranded RNA is expressed in only the target cell. The combination of cell-specific promoter and heat shock promoter drives the expression of double-stranded RNA at the optional timing by heat shock

3.3 Generation of Knockdown of Genes

1. Constructions of transgenes

   An amplified target region for RNAi was inserted downstream of the cell-specific promoter or the heat shock promoter *hsp16-2*. Transgenic animals used in T.C.RNAi experiments have two DNA fragments, one of which drives the expression of a sense strand of the target region by the cell-specific promoter and the other by the expression of an antisense strand by the heat shock promoter. Heat shock promoter *hsp16-2* was amplified with primers FP = 5′-atcaagagcatttgaatcaga-3′ and RP = 5′-ttcggtacatggaaaagtagt-3′.

2. Establishment of transgenic lines

   Cell-specific promoter::*sense* (or *antisense*), *hsp*::*antisense* (or *sense*), and a transformation marker are required for T.C.RNAi. The ratio and concentration of sense and antisense RNA strands of a target gene in injection mix depend on the target gene (*see* **Notes 2** and **3**).

3. Heat shock treatment

   For the embryonic stage, adult cuticles were dissolved in the lysis solution (NaOH:NaClO = 3:4) to gain eggs. The eggs were incubated under the heat-shocked condition at 30 °C for 8 h and then shifted to the standard condition (20 °C) and cultivated to adulthood (*see* **Note 4**).
   For the L1 larval stage, eggs were kept under the standard condition for 20 h after lysis treatment. The eggs were then shifted to the heat shock at 30 °C for 16 h and switched to the standard condition and cultivated to adulthood.
   For the adult stage, adult worms were collected with basal buffer (0.5 g/L gelatin, 5 mM KPO_4, 1 mM $CaCl_2$, and 1 mM $MgSO_4$) and transferred into a 1.5 ml tube. This tube was incubated under heat shock at 33 °C for 30 min. During this treatment, the tube was inverted every 10 min to supply air. The animals were then placed on the plate with food under the standard condition for 1 h, followed by the quantification of gene knockdown (*see* **Note 5**).

3.3.1 Knockdown of GFP

Using T.C.RNAi, the knockdown of GFP which was expressed in neurons or muscles was performed. Our protocol is described below:

1. Constructions of transgenes

   For the expression in AWC and AWB neurons, the *gcy-10* promoter [8], which was amplified with primers FP = 5′-tgggtacaacaatttctc-3′ and RP = 5′-ataattggccttctgctcaaa-3′, was used.
   For the expression in AWC neurons, the *srd-17* promoter [9], which was amplified with primers FP = 5′-ccgtctaacttctttttg -3′ and RP = 5′-tattgaattggcaaatgg-3′, was used.

For the expression in vulval muscles, we used the *myo-3* promoter which was amplified with primers FP = 5′-ttgaataaaataattttccc-3′ and RP = 5′-tggatctagtggtcgtgggt-3′.
The target region of GFP was amplified with the primers Tf = 5′-atgagtaaaggagaagaact-3′ and Tr = 5′-ctatttgtatagttcatcca-3′.

2. Establishment of transgenic lines

   Injection mix including constructs for the expression of GFP, RNA strands, and a transformation marker was prepared to generate transgenic lines. In the T.C.RNAi lines, final concentrations were 5 ng/μl GFP, 40 ng/μl *gfp*(*sense*), 40 ng/μl *gfp*(*antisense*), and 5 ng/μl *lin-44p*::GFP (*see* **Note 2**). In the control lines, final concentrations were 5 ng/μl GFP, 80 ng/μl pPD49.26, and 5 ng/μl *lin-44p*::GFP.

3. Quantification of GFP knockdown

   We monitored the intensity of GFP and examined whether expression of GFP can be decreased in the time and cell-specific manner by GFP T.C.RNAi. To clearly observe the effect of RNAi, we photobleached pre-existing GFP, which otherwise has a relatively long half life [10] (*see* **Note 6**), before T.C.RNAi and observed the recovery of GFP fluorescence after T.C.RNAi. This way we can assess the decrease of mRNA abundance caused by RNAi. GFP was photobleached by tenfold stronger excitation light than that for observation. Such photobleaching decreased the intensity of GFP to 50–70% of the intact intensity. To quantitatively compare the GFP intensity before and after T.C.RNAi, we measured the GFP intensity in the same individuals before photobleaching (intact), just after photobleaching and after a heat shock for 30 min at 33 °C and a recovery for 1 h at 20 °C at the same intensity of the excitation light. Then we calculated the rate of change in the fluorescence intensity of each animal as [(GFP intensity after heat shock) − (GFP intensity after photobleaching)]/[(intact GFP intensity) − (GFP intensity after photobleaching)]. To quantify and normalize the recovery of the intensity after heat shock treatment, the recovery ratio was calculated as [(the rate of change in the GFP intensity of each animal after the heat shock) − (the average intensity change of control treatment) + 1]. The recovery ratio of GFP intensity compared to controls (without heat shock) was measured. The intensity of GFP was measured based on the region of interest (ROI), the size of which was equivalent in all animals used for these analyses. The average intensity at three random background points was calculated as the background intensity and was subtracted from the intensity of GFP.

4. Heat shock treatment

   After photobleach of GFP, the animals were transferred to 0.2 ml tube with basal buffer (0.5 g/L gelatin, 5 mM KPO_4, 1 mM $CaCl_2$, and 1 mM $MgSO_4$). This tube was incubated under heat-shocked condition at 33 °C for 30 min or under control condition at 20 °C for 30 min, and animals in the tube were placed on the NGM plate with food at 20 °C for 1 h for the recovery.

5. GFP knockdown in neurons

   GFP was expressed by the *gcy-10* promoter which drives the expression in AWC, AWB, and I1 [8] and monitored the GFP intensity in cell bodies of AWC. In adult animals expressing both *gcy-10p*::*gfp*(*s*) and *hsp*::*gfp*(*as*), the recovery ratio of GFP intensity in AWC significantly decreased after the heat shock compared to that after the mock treatment (Fig. 2a). The expression of double-stranded RNA by the reciprocally exchanged promoter also induced GFP knockdown after the heat shock (Fig. 2a). The GFP intensity in AWC was normally recovered after the heat shock in animals without expression of the RNAi constructs (Fig. 2a). Expression of only a single RNA strand driven by a heat shock promoter or a cell-specific promoter could not decrease the recovery ratio (Fig. 2a).

   Next, GFP intensity in both AWC and AWB neurons was monitored at the same time to examine the cell specificity of the gene knockdown by T.C.RNAi. The intensity in the cell body of AWC and AWB neurons was monitored. The *gcy-10* promoter was used for the expression of GFP in both AWC and AWB neurons and the *srd-17* promoter for the expression of a single-stranded RNA specifically in AWC neurons. In animals which expressed *srd-17p*::*gfp*(*s*) and *hsp*::*gfp*(*as*), recovery of GFP intensity in AWC was reduced but not in AWB after the heat shock (Fig. 2b, c).

6. GFP knockdown in muscles

   GFP was expressed by the *myo-3* promoter which drives the expression in body muscles including vulval muscles. GFP intensity in anterior or lateral protrusions of vulval muscles was monitored. The recovery ratio of GFP intensity in vulval muscles was significantly decreased after heat shock treatment, but not after control treatment in animals with *myo-3p*::GFP, *myo-3*::*gfp*(*s*), and *hsp*::*gfp*(*as*) (Fig. 2d). GFP intensity in animals without expression of RNAi constructs was recovered normally after the heat shock (Fig. 2d).

3.3.2 Knockdown of Endogenous Genes

Knockdown of endogenous genes, *odr-3* and *let-60*, was performed.

The *odr-3* gene encodes a G protein α which mainly functions in AWC chemosensory neurons and is essential for olfactory

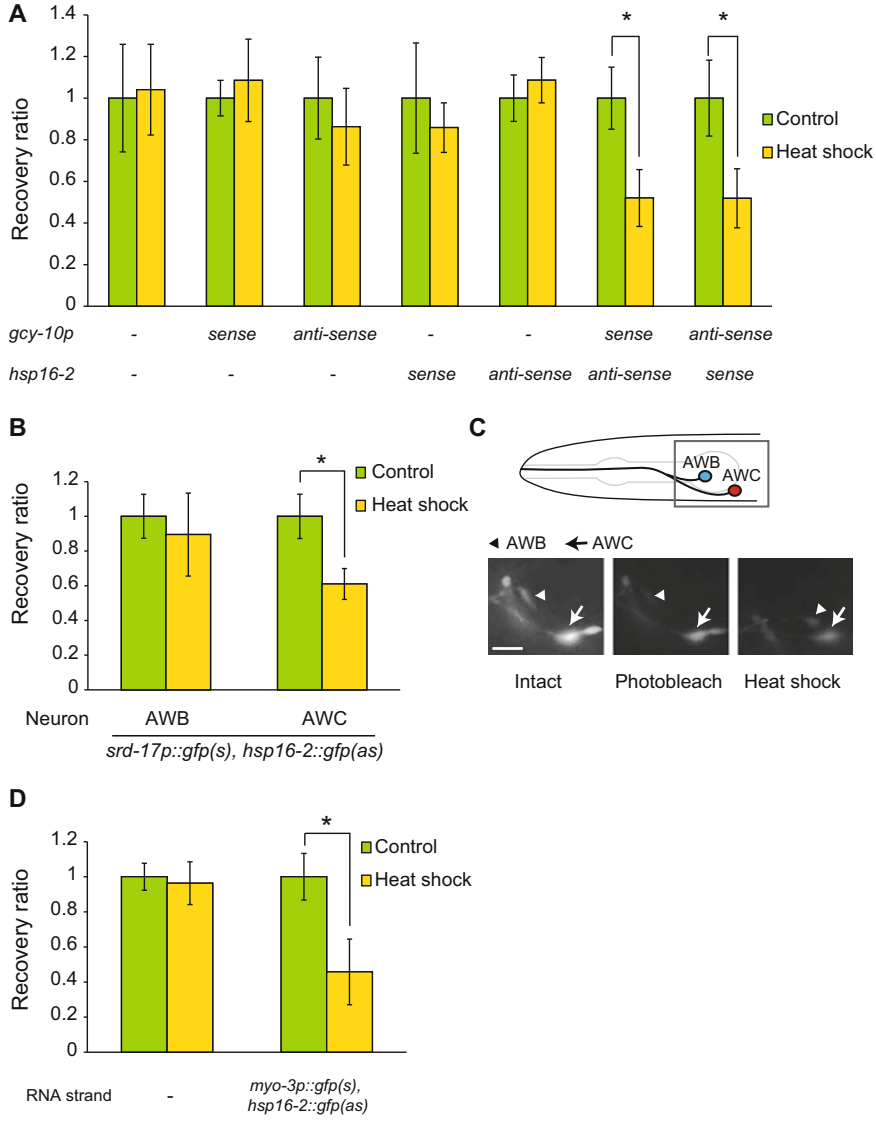

Fig. 2 Knockdown of GFP by T.C.RNAi. (**a**) The average recovery ratio of GFP intensity of AWC neurons in animals without RNAi constructs, animals with only *gcy-10p*::*gfp*(*s*), *gcy-10p*::*gfp*(*as*), *hsp16-2*::*gfp*(*s*), and *hsp16-2*::*gfp*(*as*), with both *gcy-10p*::*gfp*(*s*) and *hsp16-2*::*gfp*(*as*), or both *hsp16-2*::*gfp*(*s*) and *gcy-10p*::*gfp*(*as*) after heat-shocked or control condition ($n \geq 9$ animals). (**b**) The average recovery ratio of GFP intensity of AWB and AWC neurons in animals with *srd-17p*::*gfp*(*s*) and *hsp16-2*::*gfp*(*as*) after heat-shocked or control condition ($n \geq 10$ animals). (**c**) Representative images of *gcy-10p*::GFP in animals with *srd-17p*::*gfp*(*s*) and *hsp16-2*::*gfp*(*as*) before (*left panel*, intact) and after (*center panel*, photobleach) photobleaching and after heat-shocked condition (*right panel*, heat shock). *Arrowheads* indicate AWB neurons, and arrows indicate AWC neurons. Scale bars = 10 μm. (**d**) The average recovery ratio of GFP intensity of anterior or lateral protrusions of vulva in animals without RNAi constructs or animals with *myo-3p*::*gfp*(*s*) and *hsp16-2*::*gfp*(*as*) after heat-shocked or control condition ($n \geq 14$ protrusions of vulva). Error bars represent SEM and asterisks indicate significant differences (*$p < 0.05$, Student's *t*-test)

responses to odorants, including isoamyl alcohol, which is sensed by AWC [11]. We found the mutants of *let-60*, which encodes Ras and is involved in vulval induction, olfaction, and germ-line apoptosis [12–14], show abnormal locomotion behavior in which animals continued to move in a circled pattern, termed as circular locomotion (CL). The knockdown of *let-60* in RMD neurons, which is one of the motor neurons regulating head movements [15], induces CL. In addition, the mutants of *glr-1*, which encodes an AMAP-type glutamate receptor, and is expressed in motor neurons including RMD [15, 16], also show CL. To examine the relation between *let-60* and *glr-1* and the role of *let-60* in the control of locomotion behavior, time- and cell-specific *let-60* knockdown was performed.

Detailed conditions are described below:

1. Constructions of transgenes

   The promoter and target region used in the knockdown of odr-3 are described below.

   For the expression in AWC neurons, the gcy-10 promoter, which was amplified with same primers described in Subheading 3.3.1, was used.

   The target region of odr-3 was amplified with Tf = 5′-ctcatgccagagcaatgaaa-3′ and Tr = 5′-atgcgtttgctctctcaggt-3′.

   The promoters and the target region used in the knockdown of let-60 are described below.

   For the expression in RMD neurons, the rig-5a promoter [17], which was amplified with primers FP = 5′-attacttgtacatttcca-3′ and RP = 5′-tgatggttgttgaattg-3′, and the mgl-1 promoter [18], FP = 5′-gattttgcagaacttgga-3′ and RP = 5′-tatttcgcgatttttttc-3′, were used.

   The target region of let-60 was amplified with primers Tf = 5′-aaatccttctccacttcgttttc-3′ and Tr = 5′-aagaggatcgatcacaagtttca-3′.

   GLR-1::GFP was constructed as previously reported [19]. GFP and GLR-1::GFP were inserted downstream of the promoters.

2. Transgenic lines

   Injection mix was prepared to generate transgenic lines. In the knockdown of *odr-3* gene, final concentrations were 40 ng/μl *odr-3*(*sense*), 40 ng/μl *odr-3*(*antisense*), and 5 ng/μl *myo-3p*::GFP (*see* **Note 3**). In the knockdown of *let-60* for the behavioral analysis, final concentrations were 10 ng/μl *let-60*(*sense*), 10 ng/μl *let-60*(*antisense*), 70 ng/μl pPD49.26, and 10 ng/μl *myo-3p*::GFP. In the knockdown of *let-60* for the analysis of GLR-1 localization, final concentrations were 40 ng/μl GFL-1::GFP, 10 ng/μl *let-60*(*sense*), 10 ng/μl *let-*

60(*antisense*), 30 ng/μl pPD49.26, and 10 ng/μl *lin-44p*::GFP (*see* **Note 3**).

3. Quantification of *odr-3* knockdown

   To examine whether the response to isoamyl alcohol is decreased by T.C.RNAi, chemotaxis assays were performed [20]. In each experiment, 30–50 individuals were used. Sodium azide (1 M) is used for trapping worms. The assay format described in Fig. 3a was used. The chemotaxis index was calculated as [the number of animals in A – the number of animals in B]/[the number of animals in A, B, and C] (Fig. 3a) [20]. 10^{-3} dilution of isoamyl alcohol was used as an attractive odorant.

4. Quantification of *let-60* knockdown

   Behavioral analysis and the analysis of GLR-1 localization were performed. In behavioral analysis, whether the knockdown of *let-60* by T.C.RNAi cause CL was examined (*see* **Note 7**). In the analysis of GLR-1 localization, whether the knockdown of *let-60* by T.C.RNAi induces GLR-1 mislocalization was investigated. To analyze GLR-1 localization quantitatively, we measured the intensity and an area of puncta of GLR-1 using region of interest (ROI). The size of ROI was equivalent in all animals used for these analyses. The average intensity at three random background points was calculated as the background intensity and was subtracted from the intensity of the puncta.

5. Heat shock treatment

   Same to the condition described in Subheading 3.2, 2. Heat Shock Treatment.

6. *odr-3* knockdown

   The *gcy-10* promoter was used for the expression of an *odr-3* RNA strand. Animals in which *odr-3* was knocked down in AWC neurons by T.C.RNAi at the adult stage showed a defect in the response to isoamyl alcohol (Fig. 3b). However, the adult transgenic animals that had undergone heat shock treatment at the L1 larval stage exhibited a normal response (Fig. 3b). The adult wild-type animals that had undergone heat shock treatment at the adult stage exhibited a normal response (Fig. 3b).

7. *let-60* knockdown

   GLR-1::GFP was expressed in RMD neurons by the *mgl-1* promoter (*see* **Note 8**) (Fig. 4a), and *let-60* RNA strands were expressed by the *rig-5a* promoter (*see* **Note 9**). GLR-1 localization was not affected by *let-60* knockdown at the embryonic stage (Fig. 4b, c), while the knockdown of *let-60* at the adult stage caused significant abnormality in the localization of GLR-1 (Fig. 4b, c), which was the same phenotype induced by cell-specific RNAi.

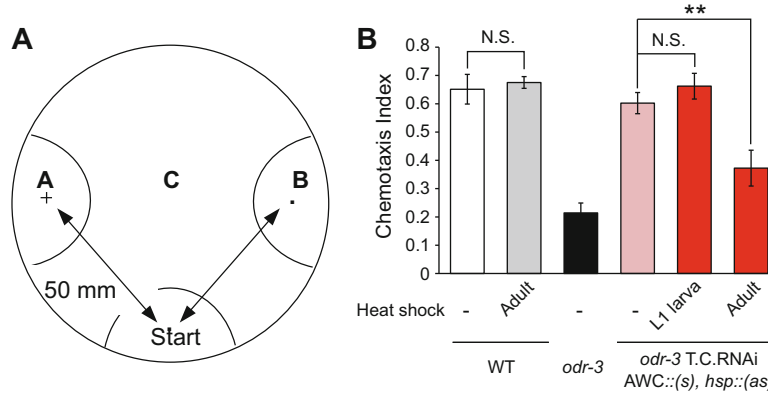

Fig. 3 Knockdown of *odr-3* by T.C.RNAi. (**a**) The plate formats for the analyses of chemotaxis. Chemotaxis index = [the number of animals in A – the number of animals in B]/[the number of animals in A, B, and C]. (**b**) Chemotaxis to isoamyl alcohol (1:1000) in wild-type animals, *odr-3* mutants, and wild-type animals with *odr-3*T.C.RNAi ($n \geq 4$ assays). AWC::(*s*) and *hsp*::(*as*) mean *gcy-10p*::*odr-3*(*s*) and *hsp16-2*::*odr-3*(*as*). Error bars represent SEM and *asterisks* indicate significant differences (**$p < 0.01$, Dunnett's test)

Next, the stage at which LET-60Ras in RMD neurons regulated locomotion behavior was investigated. The knockdown of *let-60* in RMD neurons at the L1 larval stage resulted in no defects in locomotion behavior, whereas the expression of *let-60* double-stranded RNA at the adult stage significantly ($p < 0.01$) increased the rate of CL (Fig. 4d). The same result was obtained even when the sense and the antisense strands were expressed by the reciprocally changed promoter (Fig. 4d). These results were consistent with the earlier finding that Ras in RMD neurons at the adult stage was necessary for the normal localization of GLR-1 (Fig. 4b, c) (*see* **Note 10**). Neither wild-type animals exposed to heat shock nor transgenic animals without heat shock showed CL (Fig. 4d).

4 Notes

1. It was reported that the activity of the heat shock promoter is also enhanced by the prooxidant juglone [4], suggesting the possibility that knockdown by T.C.RNAi can be induced without heat shock.
2. To perform GFP knockdown efficiently, you should increase concentrations of *gfp*(*s*) and *gfp*(*as*) in the injection mix and decrease the concentration of GFP in the injection mix due to the stability and a relatively long half life of GFP.

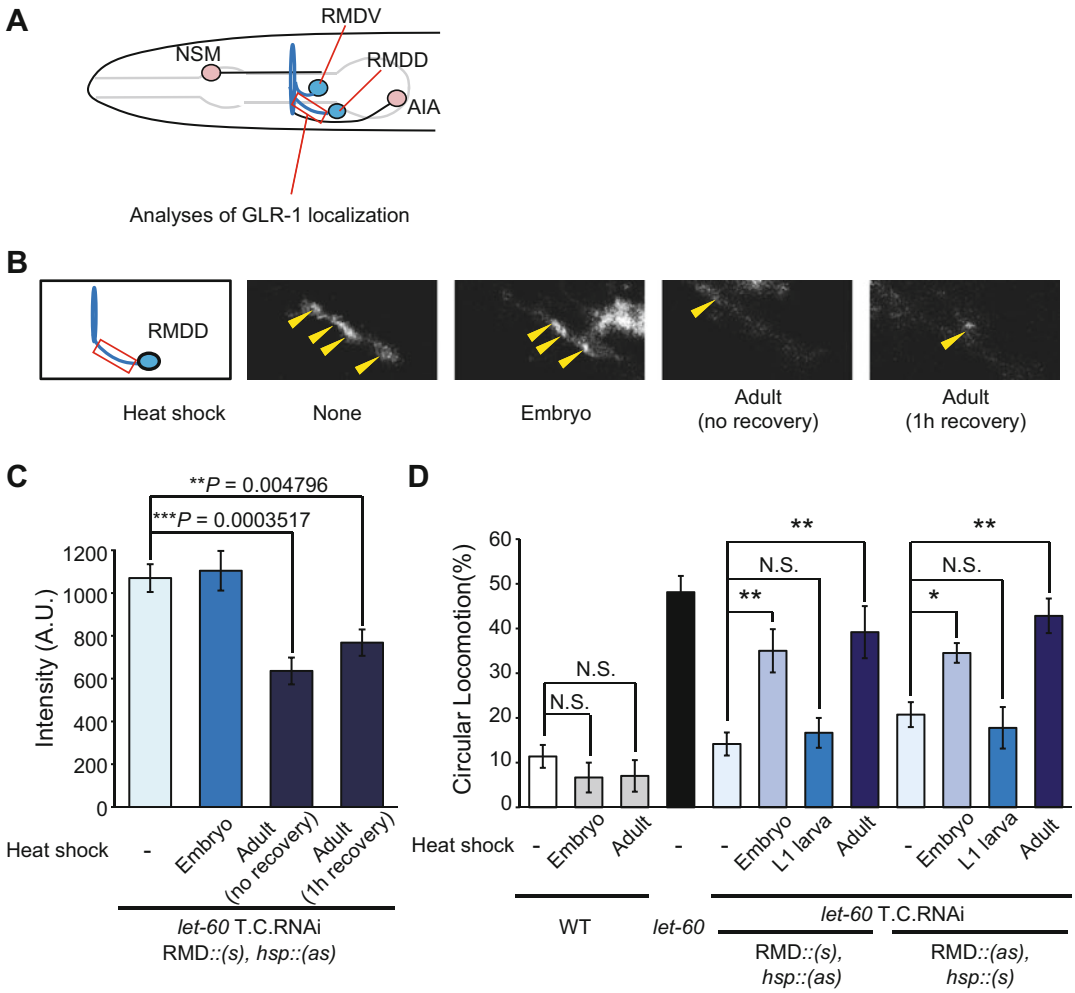

Fig. 4 Knockdown of LET-60Ras by T.C.RNAi. (**a**) *mgl-1* promoter drives expression in NSM, RMDV, RMDD, and AIA neurons. Neurites of RMDD neurons (enclosed by a *red rectangle*) were investigated by analyzing the clusters of GLR-1::GFP in RMD neurons specifically. (**b**) Representative images of GLR-1::GFP in neurites of RMDD of transgenic animals with *let-60* time- and RMD-specific RNAi (T.C.RNAi) (*rectangle* = the neurite of RMDD; *yellow arrowheads* = GLR-1::GFP clusters). (**c**) Average intensity of GLR-1::GFP fluorescence in RMDD neurons of transgenic animals with *let-60* T.C.RNAi after heat shock at the embryonic or adult stage ($n \geq 10$ animals). (**d**) The proportion of wild type, *let-60*(*lf*) mutants, and animals with *let-60* T.C.RNAi after heat shock treatment exhibiting CL ($n \geq 3$ assays). In (**c**) and (**d**), RMD::(*s*) and *hsp*::(*as*) mean *rig-5ap*::*let-60*(*s*) and *hsp16-2*::*let-60*(*as*), respectively. Error bars represent SEM and *asterisks* indicate significant differences (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, Dunnett's test)

3. The concentrations of *sense* and *antisense* RNA strands in the injection mix depend on the endogenous gene. The excessive concentration of *hsp*::*s* or *as* might lead to the unexpected expression of double-stranded RNA even when the transgenic animals are under the standard condition. Endogenous genes, especially involved in a signaling pathway, may be sensitive and vulnerable.

4. At the early embryonic stage, the heat shock treatment induces lethality in *C. elegans*.
5. Essential genes in which defects, even if minor, have great effects on the phenotype need the strict temperature control to avoid unexpected knockdown.
6. Attachment of signal peptides such as the destruction box to GFP may be helpful to analyze the knockdown efficiency.
7. The phenotype of circular locomotion was quantified using the curving rate as described in [9].
8. The *mgl-1* promoter drives the expression in NSM, AIA, and RMD neurons, and it is difficult to distinguish the neurites of RMD and AIA neurons in the nerve ring. Thus, we observed the neurites of RMDD neurons which are distant from those of AIA (Fig. 4a).
9. It was confirmed that the *rig-5a* promoter drives the RMD-specific expression at the embryonic and adult stage.
10. The knockdown of *let-60* at the embryonic stage also significantly induced CL (Fig. 4d), suggesting a role for Ras other than the regulation of GLR-1 localization at the embryonic stage to control locomotion behavior.

Acknowledgment

We thank the Caenorhabditis Genetic Center for providing the strains. This research was supported by a JSPS Grant-in-aid for Young Scientists (A), Grant-in-Aid for Scientific Research (C), Senri Life Science Foundation, Inamori Foundation, The Kurata Memorial Hitachi Science and Technology Foundation, The Japan Health Foundation, and Mishima Kaiun Memorial Foundation.

References

1. Bacaj T, Shaham S (2007) Temporal control of cell-specific transgene expression in Caenorhabditis elegans. Genetics 176(4):2651–2655. doi:10.1534/genetics.107.074369, genetics.107.074369 [pii]
2. Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods 10(8):741–743. doi:10.1038/nmeth.2532, nmeth.2532 [pii]
3. Esposito G, Di Schiavi E, Bergamasco C, Bazzicalupo P (2007) Efficient and cell specific knock-down of gene function in targeted C. elegans neurons. Gene 395(1-2):170–176. doi:10.1016/j.gene.2007.03.002, S0378-1119(07)00118-7 [pii]
4. Strayer A, Wu Z, Christen Y, Link CD, Luo Y (2003) Expression of the small heat-shock protein Hsp16-2 in Caenorhabditis elegans is suppressed by Ginkgo biloba extract EGb 761. FASEB J 17(15):2305–2307. doi:10.1096/fj.03-0376fje, 03-0376fje [pii]
5. Johnson NM, Behm CA, Trowell SC (2005) Heritable and inducible gene knockdown in C. elegans using Wormgate and the ORFeome. Gene 359:26–34. doi:10.1016/j.gene.2005.05.034, S0378-1119(05)00310-0 [pii]
6. Brenner S (1974) The genetics of Caenorhabditis elegans. Genetics 77(1):71–94
7. Mello CC, Kramer JM, Stinchcomb D, Ambros V (1991) Efficient gene transfer in C.

elegans: extrachromosomal maintenance and integration of transforming sequences. EMBO J 10(12):3959–3970

8. Yu S, Avery L, Baude E, Garbers DL (1997) Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. Proc Natl Acad Sci U S A 94(7):3384–3387
9. Hamakawa M, Uozumi T, Ueda N, Iino Y, Hirotsu T (2015) A role for Ras in inhibiting circular foraging behavior as revealed by a new method for time and cell-specific RNAi. BMC Biol 13:6. doi:10.1186/s12915-015-0114-8, s12915-015-0114-8 [pii]
10. Corish P, Tyler-Smith C (1999) Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng 12(12):1035–1040
11. Roayaie K, Crump JG, Sagasti A, Bargmann CI (1998) The G alpha protein ODR-3 mediates olfactory and nociceptive function and controls cilium morphogenesis in C. elegans olfactory neurons. Neuron 20(1):55–67, S0896-6273(00)80434-1 [pii]
12. Perrin AJ, Gunda M, Yu B, Yen K, Ito S, Forster S, Tissenbaum HA, Derry WB (2013) Noncanonical control of C. elegans germline apoptosis by the insulin/IGF-1 and Ras/MAPK signaling pathways. Cell Death Differ 20(1):97–107. doi:10.1038/cdd.2012.101, cdd2012101 [pii]
13. Sundaram MV (2006) RTK/Ras/MAPK signaling. WormBook:1–19. doi:10.1895/wormbook.1.80.1
14. Hirotsu T, Saeki S, Yamamoto M, Iino Y (2000) The Ras-MAPK pathway is important for olfaction in Caenorhabditis elegans. Nature 404(6775):289–293. doi:10.1038/35005101
15. Hart AC, Sims S, Kaplan JM (1995) Synaptic code for sensory modalities revealed by C. elegans GLR-1 glutamate receptor. Nature 378(6552):82–85. doi:10.1038/378082a0
16. Maricq AV, Peckol E, Driscoll M, Bargmann CI (1995) Mechanosensory signalling in C. elegans mediated by the GLR-1 glutamate receptor. Nature 378(6552):78–81. doi:10.1038/378078a0
17. Schwarz V, Pan J, Voltmer-Irsch S, Hutter H (2009) IgCAMs redundantly control axon navigation in Caenorhabditis elegans. Neural Dev 4:13. doi:10.1186/1749-8104-4-13, 1749-8104-4-13 [pii]
18. Greer ER, Perez CL, Van Gilst MR, Lee BH, Ashrafi K (2008) Neural and molecular dissection of a C. elegans sensory circuit that regulates fat and feeding. Cell Metab 8(2):118–131. doi:10.1016/j.cmet.2008.06.005, S1550-4131(08)00176-9 [pii]
19. Rongo C, Whitfield CW, Rodal A, Kim SK, Kaplan JM (1998) LIN-10 is a shared component of the polarized protein localization pathways in neurons and epithelia. Cell 94(6):751–759, S0092-8674(00)81734-1 [pii]
20. Bargmann CI, Hartwieg E, Horvitz HR (1993) Odorant-selective genes and neurons mediate olfaction in C. elegans. Cell 74(3):515–527, 0092-8674(93)80053-H [pii]

Chapter 7

Cell-Penetrating Peptide-Mediated Delivery of Cas9 Protein and Guide RNA for Genome Editing

Bharathi Suresh, Suresh Ramakrishna, and Hyongbum Kim

Abstract

The clustered, regularly interspaced, short palindromic repeat (CRISPR)-associated (Cas) system represents an efficient tool for genome editing. It consists of two components: the Cas9 protein and a guide RNA. To date, delivery of these two components has been achieved using either plasmid or viral vectors or direct delivery of protein and RNA. Plasmid- and virus-free direct delivery of Cas9 protein and guide RNA has several advantages over the conventional plasmid-mediated approach. Direct delivery results in shorter exposure time at the cellular level, which in turn leads to lower toxicity and fewer off-target mutations with reduced host immune responses, whereas plasmid- or viral vector-mediated delivery can result in uncontrolled integration of the vector sequence into the host genome and unwanted immune responses. Cell-penetrating peptide (CPP), a peptide that has an intrinsic ability to translocate across cell membranes, has been adopted as a means of achieving efficient Cas9 protein and guide RNA delivery. We developed a method for treating human cell lines with CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs that leads to endogenous gene disruption. Here we describe a protocol for preparing an efficient CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs, as well as treatment methods to achieve safe genome editing in human cell lines.

Key words Cas9 protein purification, Cas9 conjugation, Dialysis, In vitro sgRNA synthesis, Protein delivery, T7E1 assay

1 Introduction

The clustered, regularly interspaced, short palindromic repeat (CRISPR)-associated (Cas) system is an RNA-guided DNA cleavage system first identified in *E. coli* as a part of the adaptive immune system [1, 2]. RNA-guided DNA cleavage by Cas9 nuclease is an effective mechanism for gene silencing of foreign nucleic acids in prokaryotes and is now considered an innovative tool for targeted genome engineering. The CRISPR/Cas system has been utilized in various systems including bacteria [3], model organisms [4–12], plants [13–15], and human cell lines [16–19].

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_7,

RNA-guided endonucleases (RGENs) consist of two components: the Cas9 protein and a guide RNA. To obtain highly efficient genome editing, it is crucial to achieve successful delivery of both of these components. The conventional delivery method of Cas9 has been via plasmid, non-integrating viral carriers, such as adenoviral vectors, adeno-associated viral vectors, and non-integrating lentiviral vectors [16–21]. However, the conventional methods have some drawbacks. Plasmid- or lentiviral-mediated delivery of Cas9 into cells can result in uncontrolled integration of the plasmid sequence into the host genome leading to unwanted immune responses and the potential for safety problems caused by bacterial sequences [22, 23]. Additionally, plasmid-mediated RGEN delivery methods require tools such as transfection reagents or special instruments for microinjection or electroporation. Thus, we recently developed a simple method of treating cells to be transfected with cell-penetrating peptide (CPP)-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs that can lead to endogenous gene disruption in human cell lines (Fig. 1).

Among the programmable nucleases, zinc finger nucleases (ZFNs) are the only nucleases that have been used to deliver proteins directly into human cells [24], while transcription activator-like effector nucleases (TALENs) and RGENs each required a cell-penetrating peptide (CPP) for successful delivery into cells [25, 26]. Here, we modified the Cas9 protein and an sgRNA by conjugating and complexing them, respectively, to CPP; treatment of human cell lines with these reagents led to efficient gene disruption [26]. In this chapter we describe a set of protocols outlining an efficient method to generate CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs for safe genome editing in human cells.

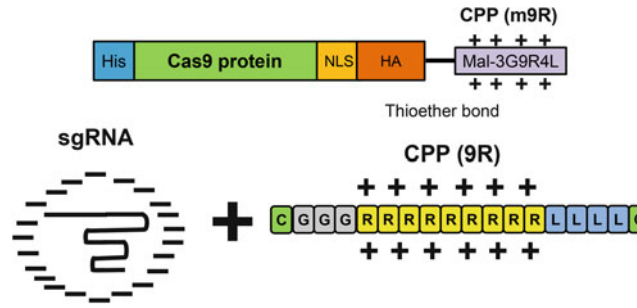

Fig. 1 Schematic representation of the Cas9 protein conjugated to a cell-penetrating peptide (CPP) and sgRNA complexed with a CPP. The single-letter codes for amino acids are used (C, cysteine; G, glycine; R, arginine; L, leucine). *His* histidine tag, *NLS* nuclear localization signal, *HA* hemagglutinin tag, *Mal* maleimide, *sgRNA* single-guide RNA

2 Materials

All solutions must be prepared using ultrapure water (with a sensitivity of 18 MΩ at 25 °C obtained using a Milli-Q Integral Water Purification System) and analytical grade reagents. Unless specified otherwise, all reagents are stored at room temperature.

2.1 Molecular Cloning Components

1. Cloning plasmids: *pET28-(a)* (T7 promoter + N-terminal 6× Histidine tag).
2. PCR amplification of SpCas9-Cys. *See* Subheading 3.1, **step 1** for an explanation regarding the additional cysteine residue at the C-terminal region of Cas9.
3. Restriction enzymes: NotI, buffer III.
4. QIAquick Gel Extraction Kit (QIAGEN).
5. Ligation reagents: 10× T4 DNA ligation buffer (NEB, New England Biolabs), T4 DNA Ligase.
6. Competent cells (DH5α, Invitrogen) and bacterial growth reagents.
7. LB agar containing 50 μg/mL kanamycin.
8. Plasmid DNA extraction kit.
9. Standard DNA gel electrophoresis reagents.

2.2 Protein Expression Components

1. Luria broth.
2. Kanamycin stock solution (50 mg/mL).
3. Isopropyl β-D-thiogalactopyranoside (IPTG).
4. Shaking incubator.

2.3 Protein Purification Components

1. Lysis buffer (20 mM Tris–HCl, 10 mM NaCl, 1 mM EDTA, pH 8.0, 1× protease inhibitor cocktail, 1 mg/mL lysozyme).
2. Sonicator.
3. Tabletop centrifuge.
4. Ni-NTA agarose resin beads.
5. Liquid chromatography protein purification system.
6. Purification buffers:
   (a) Washing buffer: 20 mM Tris–Cl, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1× protease inhibitor cocktail, 1 mg/mL lysozyme.
   (b) Elution buffer: 20 mM Tris–Cl, pH 8.0, 300 mM NaCl, 250 mM imidazole, 1× protease inhibitor cocktail.
   (c) Storage buffer: 50 mM Tris–HCl, pH 8.0, 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 20% glycerol.

7. Dialysis membrane.
8. 1× DPBS.

2.4 SDS Polyacrylamide Gel Components for Validating Protein Purity

1. Resolving gel buffer.
2. Stalking gel buffer.
3. Ammonium persulfate (10% solution in water).
4. N,N,N,N′-tetramethylethylenediamine (TEMED).
5. SDS-PAGE running buffer: 0.025 M Tris–HCl, pH 8.3, 0.192 M glycine, 0.1% SDS.
6. SDS lysis buffer (5×): 0.3 M Tris–HCl, pH 6.8, 10% SDS, 25% β-mercaptoethanol, 0.1% bromophenol blue (BPB), 45% glycerol.
7. Boiling water bath for sample boiling.

2.5 Protein Estimation Components

1. Bradford protein assay reagent.
2. 96-well plates.
3. Spectrophotometer.

2.6 RNA Synthesis Components

1. MEGAshortscript T7 transcription kit.
   (Kit components: Enzyme mix, 10× reaction buffer, nuclease-free water, TURBO DNase, ammonium acetate stop solution, ATP, CTP, GTP, and UTP solutions).
2. Template oligonucleotides (*see* Subheading 3.3, **step 3**).
3. Gel Purification Kit.
4. 8% denaturing urea-PAGE gel.
5. Phenol:chloroform, chloroform, 70% ethanol for RNA purification.
6. Spectrophotometer.

2.7 Protein–Peptide Conjugation or RNA–Peptide Complex Components

1. 4-maleimidobutyryl-GGGRRRRRRRRRLLLL peptide (m9R) for conjugation with the Cas9 protein.
2. CGGGRRRRRRRRRLLLLC peptide (9R) for formulation with in vitro synthesized sgRNA.
3. 1× DPBS.
4. Rotor.
5. Dialysis membrane.

2.8 Cell Culture Components

1. Cell line: human embryonic kidney cell line (HEK293T). (For additional cell line information, *see* **Note 1**).
2. Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS, Gibco/Invitrogen) and a penicillin/streptomycin mix (100 U/mL and 100 μg/mL, respectively) (For cell culture reagents for other cell lines, *see* **Note 2**).

3. Dissociation reagent trypsin 0.25%.
4. Opti-MEM medium.
5. 60-mm culture dishes.
6. 4-well culture dishes.
7. Hemocytometer.
8. Inverted microscope.
9. Carbon dioxide incubator.

2.9 Genomic DNA Isolation Components

1. Genomic DNA Extraction Kit
   (Kit components: nuclei lysis buffer, protein precipitation buffer, DNA resuspension buffer).
2. RNase A.
3. Water bath.
   Centrifuge.

2.10 Gene Disruption Detection and DNA Sequencing Components

1. High-fidelity DNA polymerase (*Pfu*, Promega, or Phusion polymerase).
2. Thermocycler (Bio-Rad).
3. Water bath.
4. T7E1 enzyme and NEB buffer 2 (NEB).
5. 6× DNA loading buffer.
6. Agarose for preparation of a 2% agarose gel.
7. Standard DNA gel electrophoresis reagents.
8. 100 bp DNA ladder.
9. T vector (e.g., pGEM-T Easy Vector, Promega).
10. Plasmid Isolation Kits.
11. Sequencing primers.

3 Methods

3.1 Cas9 Protein Expression Using Bacterial Expression Vector

1. A codon for an additional cysteine residue is incorporated at the C-terminus of the Cas9 gene by PCR amplification using a full-length humanized Cas9 sequence containing the HA epitope and the NLS sequence at its C-terminus as the template [16] (*see* **Note 3**).
2. The amplified Cas9 gene is then cloned into the pET28-(a), bacterial expression vector encoding a His-tag at the N-terminus driven by the T7 promoter (Fig. 2).
3. To express the Cas9 protein, a pET28-(a) vector encoding Cas9 is transformed into *E. coli* BL21 cells, followed by plating onto Luria-Bertani agar medium containing 50 μg/mL kanamycin.

Fig. 2 Structure of the bacterial expression vector, pET28a, encoding the Cas9 gene. *T7* T7 RNA polymerase promoter, *6× His tag* hexahistidine tag, *hSpCas9* human codon-optimized *Streptococcus pyogenes* Cas9, *NLS* nuclear localization signal, *HA* hemagglutinin tag

4. A single transformed *E. coli* colony is cultured in Luria broth containing 50 μg/mL kanamycin as a starter culture overnight.
5. On the following day, 0.1 OD600 of starter culture is inoculated into Luria broth containing 50 μg/mL kanamycin and incubated for 2 h until the OD600 reached between 0.4 and 0.8.
6. To standardize the Cas9 protein expression, IPTG is added to a final concentration of 0.1 mM, and the culture is incubated at 37 °C for 4 h or 0.5 mM at 30 °C for overnight.
7. After the given incubation time, protein expression is arrested by incubating the culture at 4 °C, and cells were pelleted by centrifugation (18,407 × *g*) at 4 °C.
8. Most of the expressed Cas9 proteins formed inclusion bodies at 37 °C in either 0.1 mM or 0.5 mM IPTG, while about 50% of Cas9 proteins are expressed in soluble fraction at 30 °C overnight with 0.5 mM IPTG (Fig. 3).

3.2 Cas9 Protein Purification

1. Cells are collected by centrifugation and lysed by sonication (40% duty, 10-s pulse, 30-s rest, for a total of 10 min, on ice) in a lysis buffer.
2. After lysis, bacterial cell membrane is pelleted down by centrifugation at 20,000 × *g* for 20 min at 4 °C, and the supernatant is separated and collected into a new container.
3. The Cas9 protein is purified from the soluble fraction using a column containing Ni-NTA agarose resin and an AKTA prime instrument. The entire purification process should be carried out at 4 °C.
4. The AKTA program is created to equilibrate the column, load the sample onto the column through a buffer inlet, elute the bound proteins, and regenerate the column.
5. The purification procedure consists of two sequential steps – affinity chromatography and dialysis:
   (a) At the beginning of the purification process, the column is washed with 15 column volumes of washing buffer containing 20 mM imidazole.

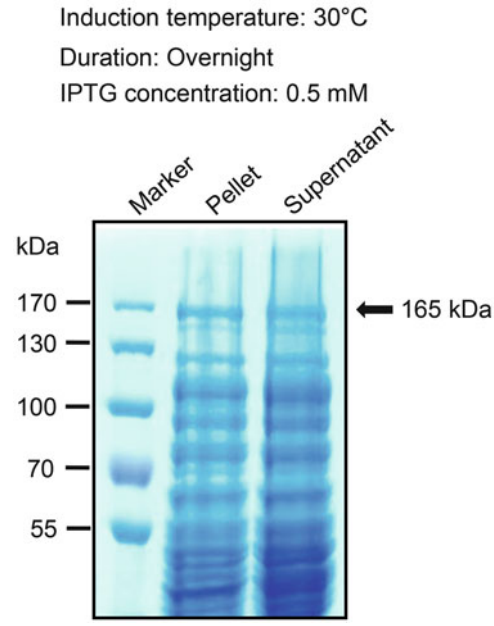

Fig. 3 Expression of His-tagged Cas9 protein in BL21 *E. coli* cells. SDS-PAGE of insoluble (pellet) and soluble fractions (supernatant) after induction of protein expression at 30 °C overnight. *Arrows* indicate the expected position of the Cas9 protein. *IPTG* isopropyl-β-D-thiogalactopyranoside

(b) During the affinity step, the lysed samples are loaded onto Ni-NTA agarose resin columns at a flow rate of 1 mL/min. Subsequently, the column is washed with 15 column volumes of washing buffer containing 20 mM imidazole at a flow rate of 1 mL/min. During the washing step, loosely attached contaminating proteins are eluted from the column.

(c) The Cas9 with His-tagged proteins are finally eluted with 2 column volumes of elution buffer containing 250 mM imidazole at a flow rate of 0.5 mL/min and collected in a fresh tube.

(d) For the dialysis step, the eluted samples are loaded into a dialysis bag and dialyzed against storage buffer (50 mM Tris–HCl, pH 8.0, 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 20% glycerol (*see* **Note 4**). The samples were aliquoted, frozen, and stored at −80 °C.

6. A sample aliquot from each step of the protein purification process is analyzed on SDS-PAGE by Coomassie blue staining (Fig. 4), and the protein concentration is determined using the Bradford assay.

3.3 sgRNA Preparation

1. RNA is transcribed in vitro through a runoff reaction using the T7 RNA polymerase and a MEGAshortscript kit.

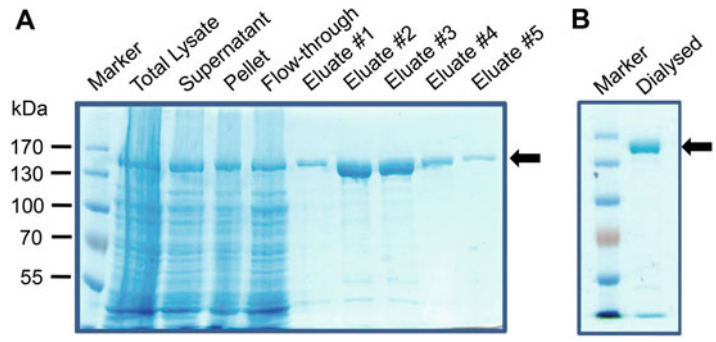

Fig. 4 SDS-PAGE of purified Cas9 protein. (**a**) The 6× His-tagged Cas9 protein was purified from the soluble fraction using a Ni-NTA column. (**b**) Pooled elution fractions 2–4 were dialyzed and used as the source of Cas9-m9R. *Arrows* indicate the expected position of the Cas9 protein

2. Templates for sgRNA transcription are generated by annealing and extension of two complementary oligonucleotides targeting our gene of interest.
3. The designed oligonucleotide template contained the T7 promoter sequence and a target sequence (5′-N20-3′).

   Forward primer (Template) 5′GAAATTAATACGACTCAC TATAGG**TGACATCAATTATTATACATG**TTTTA GAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG 3′ (The CCR5 target sequence is in bold, and the T7 promoter sequence is underlined).

   Universal reversal primer

   5′AAAAAAGCACCGACTCGGTGCCACTT TTTCAAGTTGATAACGGACTAGCCTTATTTTAACT TGC 3′
4. The frozen reagents are thawed, and the transcription reaction is assembled at room temperature. A single reaction of 20 μL is shown below. Reaction volume is scaled according to the requirement.

T7 10× Reaction Buffer	2 μL
T7 ATP Solution (75 mM)	2 μL
T7 CTP Solution (75 mM)	2 μL
T7 GTP Solution (75 mM)	2 μL
T7 UTP Solution (75 mM)	2 μL
Template DNA	<8 μL
T7 Enzyme Mix	2 μL
Nuclease-free water	Final volume to 20 μL

5. The reaction mixture is gently agitated and incubated at 37 °C for at least 4 h or overnight (*see* **Note 5**).
6. One microliter of TURBO DNase is added to the reaction mix, and it is incubated at 37 °C for another 15 min to remove any residual DNA template (*see* **Note 6**).
7. The transcribed RNA is resolved on an 8% denaturing urea-PAGE gel. RNA is recovered in nuclease-free water followed by phenol/chloroform extraction, chloroform extraction, and ethanol precipitation.
8. Purified RNA is quantified by spectrometry and stored at −80 °C.

3.4 Conjugation of m9R to the Cas9 Protein

1. Purified Cas9 protein (1 mg) in storage buffer and m9R peptide (50 μg) in PBS, pH 7.4, is prepared for conjugation (*see* **Note 7**).
2. The conjugation is carried out by dropwise mixing of m9R with Cas9 protein with constant tapping of the tube to ensure uniform mixing.
3. The reaction is allowed to proceed for 2 h at room temperature on a rotator or a vibrator.
4. Finally, the excess-free or unconjugated m9R peptides are removed by dialyzing the samples against DPBS (pH 7.4) at 4 °C for 24 h. DPBS is changed two or three times during dialysis (*see* **Note 8**).
5. For dialysis, a 50-kDa molecular weight cutoff membrane is used to ensure that only the Cas9-m9R protein (~168 kDa) is retained in the dialysis bag, while the free m9R (~2 kDa) is removed.
6. Cas9-m9R protein is collected from the dialysis membrane, and protein concentrations are determined using the Bradford assay.

3.5 Cas9-m9R Protein and sgRNA:9R Treatments

1. Cells are seeded at a concentration of 1×10^4 cells/well in a 4-well plate 24 h prior to protein treatment.
2. Twenty-four hours after plating, cells are washed with Opti-MEM (without antibiotics) (*see* **Note 9**) and incubated for 1 h prior to protein treatment.
3. The sgRNA:9R complex is formed by incubation of 10 μg of sgRNA and 30–50 μg of 9R peptide in 100 μl of Opti-MEM medium at room temperature for 30 min to form nanoparticles (Fig. 5).
4. A mixture of 150 μl Cas9-m9R (2 μM) protein and 100 μl sgRNA:9R (10:50 μg) complex is applied to the cells for 4–6 h at 37 °C.

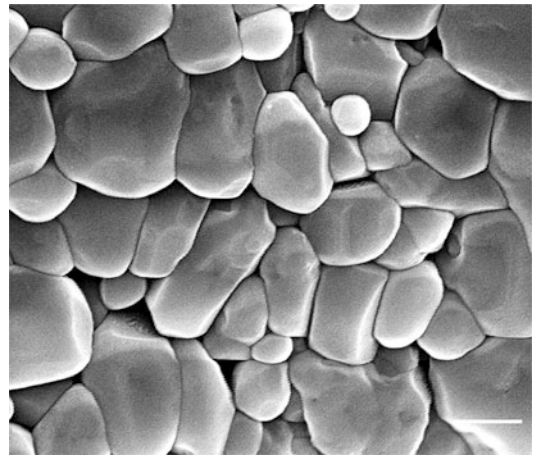

Fig. 5 Scanning electron microscopy image of nanoparticles with an sgRNA:9R weight ratio of 1:5. Scale bar = 500 nm

5. After 4 h of Cas9 protein treatment, cells are washed with PBS and incubated at 30 °C in serum-containing complete media.
6. **Steps 1–5** are repeated for two more days. Cells that had been treated for three times over three separate days are harvested by trypsinization for further analysis.

3.6 Evaluation of Mutant Cells and Sequencing

1. To determine the mutation frequency in Cas9 protein-treated cells, the genomic DNA is isolated.
2. A region containing the Cas9 target site can be PCR amplified using high-fidelity DNA polymerase (*see* **Note 10**).
3. The PCR amplicons are denatured by heating and are annealed to form heteroduplex DNA that is subjected to the T7E1 assay.
4. Five units of T7 endonuclease 1, mismatch-sensitive nucleases are incubated with self-hybridized DNA for 15–20 min at 37 °C (*see* **Note 11**).
5. T7 endonuclease 1 recognizes and cleaves heteroduplex DNA at mismatches and extrahelical loops formed by single or multiple nucleotides (Fig. 6).
6. The reaction mix is then analyzed using 2% agarose gel electrophoresis. For example, in experiments involving the CCR5 gene, we designed primers to obtain a 544-bp PCR amplicon, in which the target site lies at position 271. T7E1 treatment of the heteroduplexed DNA in the CCR5 group gave rise to two DNA bands of almost the same size (271 bp and 273 bp), which appeared as a single band after gel electrophoresis (Fig. 7).
7. Mutation frequencies are calculated as previously described based on the band intensities using ImageJ software [27].

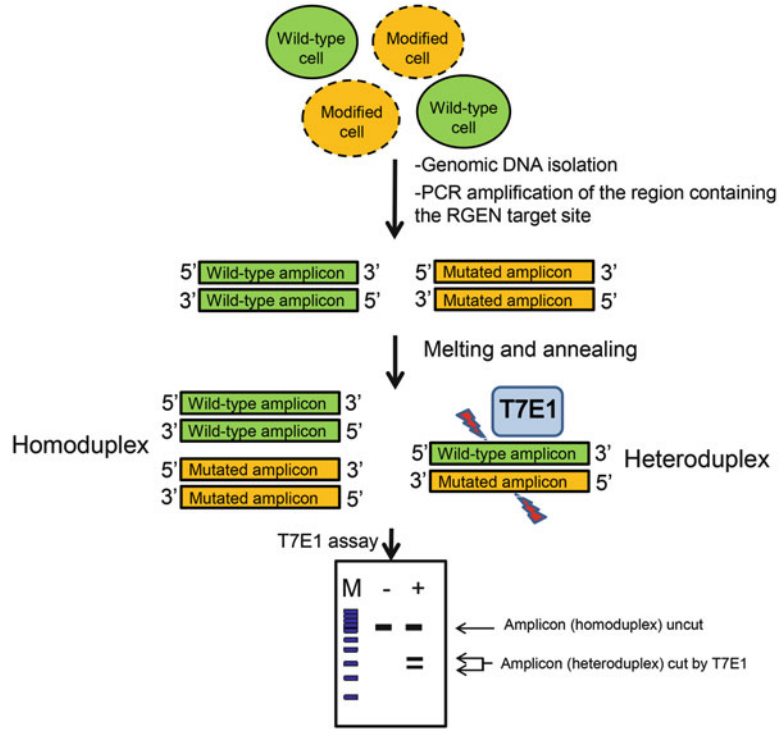

Fig. 6 Schematic overview of the T7E1 nuclease assay. DNA segments including the Cas9 target site were PCR amplified using genomic DNA isolated from cells treated with the Cas9 protein. The amplicons were melted, annealed, and subjected to T7E1 nuclease, a mismatch-sensitive endonuclease. If the amplicon contains mutated as well as wild-type sequences, heteroduplexes can be formed that can be digested with T7E1 nuclease; homoduplexes cannot be digested with T7E1 nuclease. The T7E1-treated samples were then subjected to agarose gel electrophoresis. A schematic gel result is shown. *M* size marker

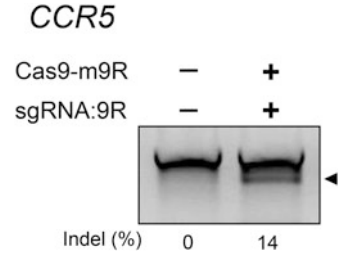

Fig. 7 The effects of CPP-mediated Cas9 delivery on endogenous gene modification. HEK293T cells were treated with Cas9-m9R and sgRNA:9R directly targeting the *CCR5* gene, and the resulting mutation frequencies were determined using the T7E1 assay. The *arrow heads* indicate the expected position of DNA bands cleaved by the T7E1 nuclease

8. The equation used to calculate the mutation frequency is $(\%) = 100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$, where the fraction cleaved is the total relative density of the cleavage bands divided by the sum of the relative density of the cleavage bands and uncut bands.
9. For sequencing, PCR products are cloned into the T vector using a pGEM-T Easy Vector PCR Cloning Kit. Cloned products are sequenced using the T7 promoter primer (5′ TAATACGACTCACTATAGGG 3′).

4 Notes

1. Other cells such as HeLa (a human cervical cancer cell line), NCCIT (a human embryonal carcinoma cell line), and H9 (a human embryonic stem cell line; from WiCell Research Institute, Madison, WI) have also been used with Cas9 protein delivery.
2. Experimental conditions and growth media may differ for each cell line. For example, 293, HeLa, and NCCIT cells are typically cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and a penicillin/streptomycin mix. The H9 cells are cultured in DMEM/F12 medium supplemented with 20% serum replacement, 1% nonessential amino acids, 1% penicillin–streptomycin, 0.1 mM β-mercaptoethanol, and 4 ng/mL basic fibroblast growth factor in the presence of mitomycin C-treated mouse embryonic fibroblasts as feeder cells.
3. The C-terminal cysteine of Cas9 containing free SH residue is capable of reacting selectively with the primary amine (-NH2) residue in m9R to form a thioether bond.
4. We observed that glycerol concentrations that were higher than 20% in the storage buffer lead to precipitation during the protein-CPP conjugation step.
5. For most of these reactions, incubation times between 2 and 4 h are sufficient. Optimum incubation time is determined based on the template size and concentration. When synthetic oligonucleotides are used as templates, overnight incubation improves RNA yield.
6. If large amount of DNA template are used, adding 2 μL of TURBO DNase helps to remove residual DNA from the reaction mix.
7. We recommend using low concentrations (about 1 mg/mL) of Cas9 protein for CPP-protein conjugation. This is the ideal concentration for efficient protein conjugation with minimal precipitation of Cas9 protein.

8. The Cas9 protein frequently precipitates during the dialysis process. The glycerol content in the storage buffer should be less than or equal to 20%, which results in minimal protein precipitation.
9. We find that using Opti-MEM improves protein delivery into cells when compared with serum-free DMEM media.
10. It is critical to obtain a single amplicon from the genomic DNA. In cases of nonspecific PCR amplicons, we recommend performing nested PCR to yield a single amplicon for reliable quantification of percent cutting efficiency.
11. Longer incubation times result in nonspecific cleavage or degradation of PCR amplicons.

Acknowledgments

This work was supported by the Korean Health Technology R&D Project, Ministry of Health and Welfare, Republic of Korea (HI14C2019 (Medistar program)).

References

1. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327(5962):167–170. doi:10.1126/science.1179555
2. Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482(7385):331–338. doi:10.1038/nature10886
3. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31(3):233–239. doi:10.1038/nbt.2508
4. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31(3):227–229. doi:10.1038/nbt.2501
5. Cho SW, Lee J, Carroll D, Kim JS (2013) Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. Genetics 195(3):1177–1180. doi:10.1534/genetics.113.155853
6. Dickinson DJ, Ward JD, Reiner DJ, Goldstein B (2013) Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. Nat Methods 10(10):1028–1034. doi:10.1038/nmeth.2641
7. Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA (2013) Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat Methods 10(8):741–743. doi:10.1038/nmeth.2532
8. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM (2013) Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029–1035. doi:10.1534/genetics.113.152710
9. Li D, Qiu Z, Shao Y, Chen Y, Guan Y, Liu M, Li Y, Gao N, Wang L, Lu X, Zhao Y (2013) Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. Nat Biotechnol 31(8):681–683. doi:10.1038/nbt.2661
10. Li W, Teng F, Li T, Zhou Q (2013) Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. Nat Biotechnol 31(8):684–686. doi:10.1038/nbt.2652
11. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153(4):910–918. doi:10.1016/j.cell.2013.04.025
12. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154(6):1370–1379. doi:10.1016/j.cell.2013.08.022

13. Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol 31(8):688–691. doi:10.1038/nbt.2654
14. Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol 31(8):691–693. doi:10.1038/nbt.2655
15. Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu JL, Gao C (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol 31(8):686–688. doi:10.1038/nbt.2650
16. Cho SW, Kim S, Kim JM, Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31(3):230–232. doi:10.1038/nbt.2507
17. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. doi:10.1126/science.1231143
18. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J (2013) RNA-programmed genome editing in human cells. Elife 2, e00471. doi:10.7554/eLife.00471
19. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826. doi:10.1126/science.1232033
20. Maggio I, Holkers M, Liu J, Janssen JM, Chen X, Goncalves MA (2014) Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. Sci Rep 4:5105. doi:10.1038/srep05105
21. Senis E, Fatouros C, Grosse S, Wiedtke E, Niopek D, Mueller AK, Borner K, Grimm D (2014) CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. Biotechnol J 9(11):1402–1412. doi:10.1002/biot.201400046
22. Hemmi H, Takeuchi O, Kawai T, Kaisho T, Sato S, Sanjo H, Matsumoto M, Hoshino K, Wagner H, Takeda K, Akira S (2000) A Toll-like receptor recognizes bacterial DNA. Nature 408(6813):740–745. doi:10.1038/35047123
23. Wagner H (2001) Toll meets bacterial CpG-DNA. Immunity 14(5):499–502
24. Gaj T, Guo J, Kato Y, Sirk SJ, Barbas CF 3rd (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat Methods 9(8):805–807. doi:10.1038/nmeth.2030
25. Liu J, Gaj T, Patterson JT, Sirk SJ, Barbas Iii CF (2014) Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. PLoS One 9(1), e85755. doi:10.1371/journal.pone.0085755
26. Ramakrishna S, Kwaku Dad AB, Beloor J, Gopalappa R, Lee SK, Kim H (2014) Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res 24(6):1020–1027. doi:10.1101/gr.171264.113
27. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ (2010) A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649:247–256. doi:10.1007/978-1-60761-753-2_15

Chapter 8

Epigenetic Analysis of Endocrine Cell Subtypes from Human Pancreatic Islets

Jia Zhang and Klaus H. Kaestner

Abstract

Epigenetic mechanisms have been proposed to contribute to multiple complex diseases, including diabetes mellitus. Dissecting the epigenomic landscape of normal and disease states has greatly expanded our understanding of the regulation of key players in disease pathogenesis and is thus opening doors to novel therapeutic avenues. The human pancreatic islet is a pivotal micro-organ that maintains global glucose homeostasis. The heterogeneity of the islet impedes meaningful in-depth epigenetic analysis using whole islet tissue, because important marks in one cell type are masked by signals from other cell types. We describe here a detailed protocol for the isolation of highly purified endocrine cell subtypes (beta, alpha, and delta cells) from human cadaveric islets, to perform cell-type-specific epigenomic analysis of histone modifications as well as global gene expression profiling.

Key words Epigenetics, FACS, Human islets, RNA extraction, ChIP-seq, Beta cell, Alpha cell, Delta cell

1 Introduction

Epigenetics, the study of heritable changes in gene function or other traits without alterations in the nucleotide sequence, is beginning to revolutionize our interpretation of disease mechanisms and therapeutic avenues [1, 2]. Epigenetic changes are considered the molecular basis that mediates the environmental impact on diabetes susceptibility, such as what can result from low birth weight, obesity, and aging [3]. Several types of epigenetic marks, for instance, various covalent modifications of the histones that from the nucleosome core, have been used to chart gene regulatory modules genome-wide. Thus, the presence of nucleosomes carrying histone H3 lysine 4 trimethylation (H3K4me3) typically marks active promoters, while histone H3 lysine 27 acetylation (H3K27Ac) is present at active enhancers, and histone H3 lysine 4 monomethylation (H3K4me1) denotes poised enhancers, for example [4]. Moreover, manipulations of epigenetic marks are able to drive desirable changes in gene expression [5, 6], opening doors to novel versatile means to delay or reserve disease progression.

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_8,

The human pancreatic islet, producing multiple potent glucose-regulating hormones, is a pivotal micro-organ that helps to maintain global glucose homeostasis. At the forefront of the pathogenesis of diabetes mellitus, a worldwide epidemic, the islet has long been in the spotlight of diabetes research. The emerging epigenetic era presents us with formidable challenges as human islets contain multiple cell types; if analyzed as a whole, the distinct epigenetic signatures of an individual cell type are masked by signals from other cell types. Adult human islets are comprised of five endocrine cell types, classified by the main hormones they produce: insulin-secreting beta cells, glucagon-producing alpha cells, somatostatin-releasing delta cells, pancreatic polypeptide-secreting PP cells, and ghrelin-releasing epsilon cells. Among these cells, beta, alpha and delta cells make up the majority of the endocrine mass, 54%, 35%, and 11%, respectively [7]. In-depth epigenetic analysis of individual endocrine subtype is required to understand normal islet function and various stages of diabetes progression and to develop mechanism-based therapeutic strategies. In line with this notion, our group discovered that human alpha cells exhibit an intriguing epigenomic plasticity, in that many beta-cell signature genes are marked by both activating and repressing histone modifications in alpha cells. In addition, once repressive marks are erased, alpha cells can reactivate the "beta-cell program," leading to ectopic expression of beta-cell genes, such as insulin and Pdx1 [6].

Here, we describe a protocol for cell-type-specific epigenomic analysis starting from isolated human islets from deceased organ donors. Briefly, as summarized in Fig. 1, we detail two fluorescence-activated cell sorting (FACS) strategies to obtain highly pure beta,

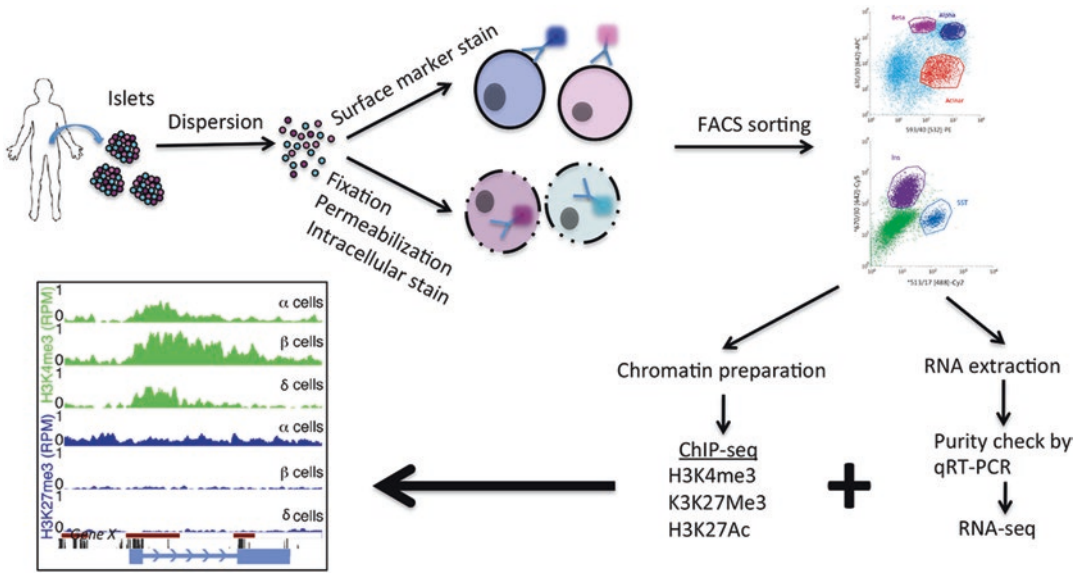

Fig. 1 Protocol outline. Overall outline of the protocol from human islets to data acquisition

alpha, and delta cells, after which individual cell populations are subjected to chromatin preparation or RNA extraction. These methods have been adapted and modified from prior publications [8, 9]. Finally, RNA-seq and chromatin immunoprecipitation following sequencing (ChIP-seq) of key histone modifications are performed. Data from these studies are then analyzed and integrated into high-resolution epigenomic maps.

2 Materials

2.1 FACS of Endocrine Cells from Human Cadaveric Islets

2.1.1 Surface Marker Sorting

1. Human cadaveric islets are obtained from the Integrated Islet Distribution Program (https://iidp.coh.org/secure/isletavail/home.aspx).
2. 37 °C water bath.
3. Eppendorf centrifuge 5804R (for 50 ml or 15 ml conical tubes).
4. 40 μM nylon strainer.
5. 1× PBS.
6. 100% fetal bovine serum (FBS).
7. 0.05% trypsin–EDTA.
8. Antibodies:
   Primary (gifts from Dr. Markus Grompe at the Oregon Health & Science University [8]).
   (a) HIC1-2B4 (HPi1) is a mouse IgG1 that labels all human islet cells (to slightly varying degrees; beta cells are a bit brighter than other endocrine cells). 1: 50 dilution.
   (b) HIC3-2D12 (HPa3) is a mouse IgM that differentially labels endocrine subtypes. Alpha cells are brightly labeled; delta cells are moderately labeled; and beta cells are dim to negative. This antibody also dimly labels duct cells, but these can be easily distinguished by their HIC1-2B4 negativity. 1:50 dilution.
   (c) HIC1-1C10 (HPx2) is a mouse IgM that labels acinar cells. 1:50 dilution.
   Secondary:
   (a) R-Phycoerythrin-AffiniPure F(ab′)2 Fragment Goat Anti-Mouse IgM, μ Chain (Jackson ImmunoResearch #115-116-075) at 1:200 dilution.
   (b) Allophycocyanin-AffiniPure Goat Anti-Mouse IgG (subclasses 1+2a+2b+3), Fcγ (Jackson ImmunoResearch #115-135-164) at 1:200 dilution.
9. FACS tubes.
10. DAPI (4′,6-diamidino-2-phenylindole, dihydrochloride).

2.1.2 Intracellular Marker Sorting

1. RNase-free microcentrifuge tubes.
2. RNaseZap.
3. 10×PBS, or 1× DEPC-treated PBS without Ca^{2+} and Mg^{2+}.
4. Bovine serum albumin (BSA).
   To make 10% BSA stock, dissolve 5 g BSA powder in 50 ml 1×PBS and filter through a sterile centrifuge tube top filter unit.
5. Saponin, molecular biology grade.
   To make 1% saponin stock, dissolve 0.5 g saponin in 50 ml 1×PBS and filter through a sterile centrifuge tube top filter unit.
6. Formaldehyde (32% solution).
   To make 1% formaldehyde, dilute with 1×PBS (nuclease-free) in a chemical hood.
7. RNasin Plus RNase Inhibitor.
8. Antibodies:
   (a) Primary:

   Guinea pig anti-insulin antibody (Invitrogen #18-0067): 1:1000.

   Goat antisomatostatin antibody (Santa Cruz #sc-7819): 1:2000.

   (b) Secondary:

   Cy5-conjugated Donkey anti-Guinea Pig IgG (Jackson ImmunoResearch #706-157-148).

   1:500.

   Cy2-conjugated Donkey anti-Goat IgG (Jackson ImmunoResearch # 705-225-147).

   1:500.
9. Low retention tubes.
10. Nanodrop 1000 spectrophotometer.
11. Qubit 3.0 Fluorometer (optional).

2.2 Sample Purity Assessment by Quantitative Real-Time PCR (qRT-PCR) and RNA-seq

1. Qiagen RNeasy Mini (#74104) or Micro (# 74004) Kit.
2. QIAshredder (Qiagen #79654).
3. 100% ethanol.
4. RecoverAll Total Nucleic Acid Isolation Kit.
5. Agilent 2100 Bioanalyzer.
6. Agilent RNA 6000 Nano Kit and Reagent.

2.2.1 RNA Extraction

2.2.2 cDNA Synthesis

1. Ovation RNA-seq System V2 (NuGEN #7102-08).
2. SuperScript II Reverse Transcriptase (Life Technologies #18064-014).

3. Oligo$(dT)_{12\text{-}18}$ (Life Technologies #18418-012).
4. dNTP (Life Technologies #10297-018).
5. RNaseOUT (Life Technologies #10777-019).

2.2.3 RNA-seq Library Preparation

1. NEBNext Ultra RNA Library Prep Kit for Illumina.
2. NEBNext PolyA mRNA magnetic isolation module.
3. NEBNext Muliplex Oligos for Illumina (Index primers set 1).
4. Illumina HiSeq 2500 System (Illumina #SY-401-9001DOC).
5. Agilent High Sensitivity DNA kit and reagent.
6. Agilent Polypropylene 96-well PCR tube plates and Optical strip caps.
7. Agilent Mx3000P qPCR system.
8. DynaMag-2 Magnet.

2.3 ChIP and ChIP-seq

1. 2.5 M glycine: dissolve 190 mg of glycine powder in 1 ml H_2O; make fresh for each use.
2. ChIP whole cell lysis buffer: 10 mM Tris–HCl, pH 8.0, 10 mM NaCl, 3 mM MgCl2, 1% NP-40, 1% SDS, 0.5% DOC.

1 M Tris–HCl (pH 8.0)	10 µl
5 M NaCl	2 µl
1 M $MgCl_2$	3 µl
10% NP-40	100 µl
10% SDS	100 µl
10% DOC	50 µl
50× protease inhibitor	20 µl
100× phosphatase inhibitor	10 µl
H_2O	705 µl
Total	1 ml

3. ChIP dilution buffer: 16.7 mM Tris–HCl pH 8.0, 167 mM NaCl, 0.01% SDS, 1.1% Triton X-100.

1 M Tris–HCl, pH 8.1	835 µl
5 M NaCl	1.67 ml
10% SDS	50 µl
10% Triton X-100	5.5 ml
H_2O	41.9 ml
Total	50 ml

4. Wash buffer TSEI: 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100.

1 M Tris–HCl, pH 8.1	1 ml
5 M NaCl	1.5 ml
0.5 M EDTA	200 µl
10% SDS	500 µl
10% Triton X-100	5 ml
H_2O	41.8 ml
Total	50 ml

5. Wash buffer TSEII: 20 mM Tris–HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100.

1 M Tris–HCl, pH 8.1	1 ml
5 M NaCl	5 ml
0.5 M EDTA	200 µl
10% SDS	500 µl
10% Triton X-100	5 ml
H_2O	38.3 ml
Total	50 ml

6. ChIP buffer III: 10 mM Tris–HCl (pH 8.0), 0.25 M LiCl, 1 mM EDTA, 1% NP-40, 1% deoxycholate.

1 M Tris–HCl, pH 8.1	500 µl
5 M LiCl	2.5 ml
0.5 M EDTA	100 µl
10% NP-40	5 ml
10% deoxycholic acid	5 ml
H_2O	36.9 ml
Total	50 ml

7. TE: 10 mM Tris–HCl (pH 8.0), 1 mM EDTA.

1 M Tris–HCl, pH 8.1	500 µl
0.5 M EDTA	100 µl
H_2O	49.4 ml
Total	50 ml

8. Elution buffer: 1% SDS, 0.1 M $NaHCO_3$.

10% SDS	100 µl
1 M $NaHCO_3$	100 µl (84 mg in 1 ml H_2O made fresh)
H_2O	800 µl
Total	1 ml

9. 1 M Tris–HCl (pH 7.5), 500 mM EDTA, 10 mg/ml Proteinase K.
10. QIAquick PCR Purification Kit.
11. Covaris M220 focused-ultrasonicator instrument.
12. microTUBE with AFA Fiber Snap-Cap.
13. Agilent DNA 1000 kit and reagent.
14. Agilent High Sensitivity DNA Kit and reagent.
15. Antibodies: H3K4me3 (Abcam #8580), H3K27Ac (Abcam # ab4729), H3K27me3 (Upstate #07-449).
16. Protein-G agarose beads.
17. BSA, molecular biology grade.
18. NEBNext ChIP-seq Library Prep Reagent Set for Illumina (#E6200S).

3 Methods

3.1 FACS of Endocrine Cells from Human Cadeveric Islets (See Note 1*)*

3.1.1 FACS of Beta and Alpha Cells from Human Islets Using Cell Surface Markers (Modified After [8])

1. To sediment islets, combine 30,000–40,000 islet equivalents (IEQ) of fresh human islets from cadaveric donors into two 50 ml conical tubes, and centrifuge 4 min at 300 × *g* at room temperature. Remove supernatant and use supernatant to rinse residual islets out of flask. Centrifuge again at 300 × *g* at room temperature for 4 min, and carefully aspirate supernatant without disturbing pellets.
2. To prepare a single-cell suspension, add 3 ml 0.05% trypsin–EDTA to each 50 ml tube; pipette up and down a few times to mix (*see* **Note 2**). Incubate in 37 °C water bath for 9 min. Pipette up and down a few times every 3 min.
3. Remove tubes from water bath. Passage contents of both tubes through one strainer into one 50-ml conical tube in the following order (*see* **Note 3**).
   (a) Passage contents of tube 1 through the strainer.

(b) Use 1 ml 100% FBS to rinse tube 1 and passage through strainer.

(c) Passage contents of tube 2 through strainer.

(d) Use 1 ml 100% FBS to rinse tube 2 and passage through strainer.

(e) Swirl tube, and if clumps develop, vortex tube for 2 s.

4. Going forward, cells remain in one conical tube. Centrifuge at 300 × *g* for 4 min, discard supernatant, and resuspend pellet in 25 ml 1×PBS. Centrifuge 4 min at 300 × *g* again and resuspend pellet in 1–2 ml 2% FBS in 1×PBS.
5. Remove 10 μl of cell suspension to determine cell number. Add appropriate amount of 2% FBS to achieve a final concentration of 5×10^6 cells/ml.
6. Set aside 100,000 cells in a FACS tube as negative control. Add appropriate amount of 2% FBS for a final volume of 500 μl. Store on ice until ready to proceed (*see* **Note 4**).
7. Add three primary antibodies HIC1-2B4, HIC3-2D12, and HIC1-1C10 at 1:50 dilution to cell suspension; incubate for 30 min at 4 °C with gentle agitation.
8. To remove unbound primary antibodies, sediment cells by centrifugation for 4 min at 300 × *g*; resuspend in 25 ml 1×PBS. Repeat centrifugation and resuspend cell pellet in 2% FBS to a concentration of 5×10^6 cells/ml.
9. Add two secondary antibodies to cell suspension at 1:200 dilution; incubate for 30 min at 4 °C with gentle agitation (*see* **Note 5**).
10. Remove unbound secondary antibodies as described in **step 8**. Transfer cell suspensions to FACS tubes.
11. To collect cells after sorting, coat new FACS tubes with 500 μl cold 1×PBS on ice. Add DAPI (final concentration: 500–1000 ng/ml) to cell suspension just prior to sorting as marker of cell viability.
12. Analyze unstained negative control first in the cell sorter to set up gating parameters (*see* **Note 6**). Separate stained cells in the sorter and collect populations according to their fluorescent signals. A typical sorting outcome is presented in Fig. 2a. Once sorted, cells are ready to be used in RNA extraction (detailed in Subheading 3.1.3) or chromatin preparation for ChIP and ChIP-seq (detailed in Subheading 3.2). Cells can be stored on ice for 1 or 2 h prior to downstream applications.

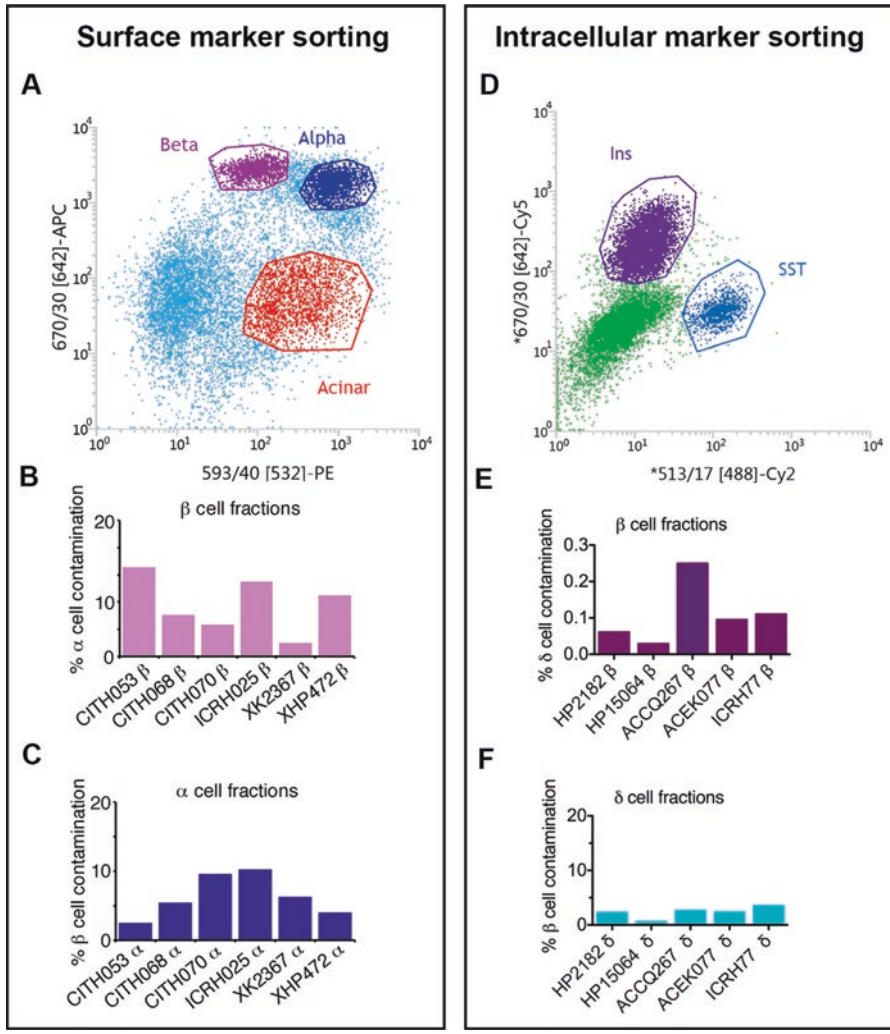

Fig. 2 Fluorescence-activated cell sorting for highly purified endocrine subtypes from human islets. (**a**, **d**) show representative sorting plots. After sorting, cells are subjected to contamination assessment in which insulin and glucagon or somatostatin gene expression levels from each fraction are measured by qRT-PCR and used to calculate the percentage of contamination (**b**, **c**, **e**, **f**). Both sorting methods provide high sample purity. Intracellular marker sorting is even more stringent, with an average 0.1 % contamination of delta cells in sorted beta cells, and 2.3 % contamination of beta cells in sorted delta cells

3.1.2 FACS of Endocrine Beta and Delta Cells from Human Islets Using Intracellular Markers (Adapted from [9])

1. Wipe clean work surfaces, centrifuge, micropipettors, Pipet-Aid, and gloves with RNaseZap.
2. Prepare the following solutions in RNase-free 1×PBS. Preincubate solutions, especially antibody solutions, with RNasin for at least 10 min prior to use to inactivate any remaining RNases. Invert gently to avoid introducing bubbles.
   (a) Fix buffer: 1 % formaldehyde (w/v), 0.1 % saponin (w/v), 1:50 RNasin.

Total volume: 1 ml/10^7 cells.

(b) Wash buffer: 0.2% BSA (w/v), 0.1% saponin, 1:200 RNasin.
Total volume: 8 ml/10^7 cells.

(c) Staining buffer: 1% BSA, 0.1% saponin, 1:25 RNasin, antibody.
Total volume: 1 ml/10^7 cells per antibody stain.

(d) Sort buffer: 0.5% BSA, 1:25 RNasin.
Total volume: 1 ml/10^7 cells.

3. Starting from as few as 16,000 IEQ human islets, generate a single-cell suspension as described in Subheading 3.1.1, **steps 1–5**.
   For quality control of sorting purity, retain some islet cells prior to fixation and extract RNA.
4. Wash cell pellet again and conduct following steps at 4 °C.
5. Resuspend pellets in "fix buffer" at 1 ml/10^7 cells. Incubate for 10–15 min at room temperature (*see* **Note 7**).
6. Centrifuge at 3000 × *g* for 3 min (*see* **Note 8**).
7. Wash cells in "wash buffer." Remove a small aliquot as staining negative control. Wash again. Resuspend the pellet in "stain buffer" at 1 ml/10^7 cells with primary antibody or no antibody (negative gating control). Incubate for 30 min while gently rocking.
8. Wash cells twice in "wash buffer" and then resuspend in "staining buffer" with secondary antibodies. Wrap tubes in aluminum foil to prevent quenching of the fluorophores. Incubate for 30 min while gently rocking.
9. Wash cells twice in "wash buffer" before resuspending in "sort buffer" (0.5–1 ml/10^7 cells). Keep cells on ice for sorting.
10. Prime and flush fluidics of cell sorter with DEPC-treated water and DEPC-treated PBS in sheath fluid. Coat the inside of FASC tubes with cold "sort buffer" for cell collection after sorting (*see* **Note 9**). A typical sorting result is shown in Fig. 2d.

3.1.3 Assessment of Sample Purity by Quantitative Real-Time PCR (qRT-PCR) and RNA-seq

RNA Extraction from Sorted Cells Using Surface Markers

Depending on the yield from FACS, use Qiagen RNeasy Mini (for 10^6–10^7 cells) or RNeasy Micro Kit (for $<5 \times 10^5$ cells) or equivalent kits from other vendors for RNA extraction.

1. Sediment sorted cells by a 4 min spin at 300 × *g*. Loosen pellet by flicking, and add RLT plus buffer supplemented with β-mercaptoethanol (10 μl per 1 ml buffer RLT plus): for $<5 \times 10^6$ cells, add 350 μl buffer; for 5×10^6 to 1×10^7 cells, add 600 μl buffer.
2. To homogenize lysate, transfer the lysate directly into a QIAshredder spin column, and centrifuge for 2 min at maximum speed (*see* **Note 10**).

3. Add 1 volume of 70% ethanol to the lysate, and mix well by pipetting. Do not centrifuge. Transfer the entire mixture to an RNeasy spin column placed in a 2 ml collection tube. Close the lid gently, and centrifuge for 15 s at 9,391 × *g*. Discard the flow through.
4. To wash the spin column membrane, add 700 µl buffer RW1 to the column and repeat the centrifugation as **step 3**; subsequently add 500 µl Buffer RPE and centrifuge for 2 min at 9,391 × *g*.
5. Place the RNeasy spin column in a new 1.5 ml collection tube. Add 30–50 µl RNase-free water directly to the spin column membrane. Close the lid gently, and centrifuge for 1 min at 9,391 × *g* to elute the RNA.
6. Repeat elution one more time with eluate to obtain a higher RNA concentration.

RNA Extraction from Fixed and Sorted Cells Using Intracellular Markers

1. Centrifuge sorted cells at 3000 × *g* for 6 min at 4 °C, and aspirate supernatant as much as possible with care not to disturb cell pellet.
2. Add 100 µl of digestion buffer and 4 µl protease from the RecoverAll Kit. Mix gently by pipetting up and down. Incubate on clean heat block at 50 °C for 3 h.
3. Proceed with the reminder of the RecoverAll procedure for RNA isolation or freeze samples at –80 °C until ready for processing. Elute the RNA in 60 µl of nuclease-free water (*see* **Note 11**).
4. Measure RNA concentration with NanoDrop 1000 or Qubit 3.0 Fluorometer. Assess RNA quality with Agilent 2100 Bioanalyzer and Agilent RNA Nano or Pico Kit. A RNA integrity number (RIN) greater than 8 is considered good quality. Store RNA at –80 °C.

qRT-PCR to Assess Sample Purity (See Representative Sample Purities from Five to Six Individual Human Donors in Fig. 2b, c, e, f)

1. cDNA synthesis
   (a) Starting material: high quality RNA, >10 ng
   Synthesize first-strand cDNA as below:
   - Add the following components to a nuclease-free microcentrifuge tube (*see* **Note 12**):

Oligo $(dT)_{12\text{-}18}$ (500 µg/ml)	1 µl
Total RNA (10–30 ng)	*X* µl
dNTP Mix (10 mM each)	1 µl
Nuclease-free, distilled water	to 12 µl

   - Incubate mixture at 65 °C for 5 min and quickly chill on ice. Briefly spin down the contents of the tube and add:

5× first-strand buffer	4 μl
0.1 M DTT	2 μl
RNaseOUT (40 units/μl)	1 μl

- Pipette to mix the content in the tube gently and incubate the tube at 42 °C for 2 min.
- Add 1 μl (200 unites) of SuperScript II RT, and mix by pipetting gently up and down and incubate at 42 °C for 50 min.
- Inactivate the reaction by 15 min-incubation at 70 °C. The cDNA is ready to be used as template for qRT-PCR.

(b) Starting material: poor-quality and low-abundance RNA (RIN < 8, amount <10 ng)
Synthesize amplified cDNA using Ovation RNA-seq V2 System (*see* **Note 13**) following the manufactures' instruction. The resulting SPIA cDNA is suitable for qRT-PCR and RNA-seq.

2. qRT-PCR

(a) Set up qRT-PCR reactions in triplicates for each sample as below:

cDNA	0.5 μl
2× SYBR Green	9.5 μl
Primer mix (forward and reverse primers, 10 μM each)	0.5 μl
H_2O	9.5 μl
Total	20 μl

Primer sequences:

Gene	Forward primer sequence	Reverse primer sequence
Hypoxanthine-guanine phosphoribosyltransferase 1 (HPRT1)	gcagactttgctttccttgg	aacacttcgtggggtccttt
Insulin (INS)	aggccatcaagcagatcact	gcacaggtgttggttcacaa
Glucagon (GCG)	gaattcattgcttggctggt	ctggcggcaagattatcaag
Somatostatin (SST)	ccaaccagacggagaatgat	ccatagccgggtttgagtta

(b) Perform the following PCR program on a qPCR instrument (e.g., Agilent Mx3000P qPCR system).

95 °C for 3 min, 1 cycle

95 °C for 5 s, 60 °C for 20 s, 40 cycles

95 °C for 1 min, 60 °C for 30 s, 95 °C for 30 s, 1 cycle

$$\text{Gene expression units} = 2^{(\text{HPRT1C}_\text{T} - \text{Gene C}_\text{T})}$$

Sample purity is calculated as percentage of contamination by the opposite cell type:

$$\%\text{Alpha cell contamination}\left(\text{in Beta cell population}\right) = \text{purity}\left(\text{Alpha}\right) \times \textit{GCG}\,\text{expression units}\left(\text{Beta cells}\right) / \textit{GCG}\,\text{expression units}\left(\text{Alpha cells}\right)$$

$$\%\text{Beta cell contamination}\left(\text{in Alpha cell population}\right) = \text{purity}\left(\text{Beta}\right) \times \textit{INS}\,\text{expression units}\left(\text{Alpha cells}\right) / \textit{INS}\,\text{expression units}\left(\text{Beta cells}\right)$$

$$\%\text{Beta cell contamination}\left(\text{in Delta cell population}\right) = \text{purity}\left(\text{Beta}\right) \times \textit{INS}\,\text{expression units}\left(\text{Delta cells}\right) / \textit{INS}\,\text{expression units}\left(\text{Beta cells}\right)$$

$$\%\text{Delta cell contamination}\left(\text{in Beta cell population}\right) = \text{purity}\left(\text{Delta}\right) \times \textit{SST}\,\text{expression units}\left(\text{Beta cells}\right) / \textit{SST}\,\text{expression units}\left(\text{Delta cells}\right)$$

Preparation of RNA-seq Library

For High-Quality, High-Abundance RNA

1. Perform mRNA isolation, fragmentation, and priming using NEBNext PolyA mRNA magnetic isolation module,
2. Using NEBNext Ultra RNA Library Prep Kit, generate RNA-seq libraries from RNA from the previous step as follows: cDNA synthesis, end repair, adaptor ligation, PCR library enrichment multiplexing, and purify PCR reaction using Agencourt AMPure XP beads.
3. Assess library quality by analyzing 1 μl of library on a Bioanalyzer using DNA high-sensitivity chip. A successful library displays a narrow distribution on the electrorpherogram, with a peak size approximately 300 bp (*see* **Note 14**).
4. Sequence libraries on Illumina HiSeq2500.

3.1.3.0.1 For Low-Integrity and/or Abundance RNA

1. Synthesize double-stranded SIPA cDNA using Ovation RNA-seq System V2.
2. To fragment cDNA to 150 bp, add 110 μl TE to 20 μl purified SIPA cDNA, and transfer total 130 μl mix to a 130 μl microTUBE with AFA Fiber snap-cap. Place microTUBE in Covaris Ultrasonicator M220 and run preloaded protocol for DNA shearing for size of 150 bp, and evaluate the outcome by DNA 1000 chip on a Bioanalyzer.

3. Following the manual of NEBNext Ultra RNA Library Prep Kit, apply 55.5 μl of fragmented cDNA from previous step to end repair, adaptor ligation, and PCR library enrichment and multiplexing, and purify PCR reaction using Agencourt AMPure XP beads.
4. Evaluate library quality and subject samples to sequencing as in subheading "For High-Quality, High-Abundance RNA" **steps 3** and **4**.

3.2 ChIP-seq Analysis of Histone Modifications

3.2.1 Preparation of Chromatin

1. Resuspend fixed and sorted cells from Subheading 3.1.2 in 130 μl cold ChIP Whole Cell Lysis Buffer.
2. Shear chromatin using Covaris Ultrasonicator M220 with the following protocol:

Duty cycle, 5%; bath temperature, 7 °C; intensity peak incident power, 75 W; cycles per burst, 200; processing time, 4 min; cell number, 0.3–3 × 10^6 cells/130 μl volume.

3. Take 10 μl sheared chromatin to prepare input and to assess chromatin shearing, add 90 μl 1×PBS and 3.5 μl NaCl, and incubate at 65 °C overnight. Snap freeze the remainder, and store at −80 °C.
4. On the next day, add 4 μl 1 M Tris–HCl (pH 7.5), 2 μl 500 mM EDTA, and 1 μl 10 mg/ml Proteinase K to input to reverse cross-linking; incubate for 1 h at 45 °C, and purify using the QIAquick PCR Purification Kit. Elute in 50 μl EB.
5. Analyze 1 μl of elute on a Bioanalyzer to check for appropriate shearing using DNA 1000 chip.

3.2.2 Immuno precipitation and Recovery of DNA

1. Thaw chromatin and place on ice. Add chromatin from minimum 10^5 sorted cells to 1 ml of ChIP dilution buffer freshly supplemented with 20 μl 50× protease inhibitor, mix with antibody (~2 μg), and rotate in cold room overnight.
2. Wash protein-G agarose with 1 ml cold ChIP dilution buffer as follows:
   (a) Add 1 ml ChIP dilution buffer to appropriate amount of agarose slurry (40 μl per sample)
   (b) Mix by inverting, then sediment the beads by a spin at 376 × *g* for 30 s, and repeat **steps a** and **b** two additional times
3. To block protein-G agarose, resuspend agarose in appropriate amount of cold ChIP dilution buffer with BSA as follows, and rotate in cold room overnight.

   Prepare 100 μl ChIP dilution buffer/BSA mix for each sample as below:

Protein-G agarose slurry	20 µl (from 40 µL of agarose slurry)
10 mg/ml BSA	10 µl
50× protease inhibitor	2 µl
ChIP dilution buffer	68 µl

4. On the next day, add 100 µl of blocked protein-G agarose to each chromatin sample, and rotate in cold room for 1 h
5. Centrifuge agarose for 30 s at 376 × *g*, and aspirate supernatant.
6. To wash protein-G agarose after immunoprecipitation, add 1 ml of room temperature buffer TSEI to agarose pellet, and rotate mix at room temperature for 5 min; sediment agarose 30 s at 376 × *g*; aspirate supernatant.
7. Repeat wash as in **step 6** using buffer TSE II, ChIP Buffer III and TE.
8. Add 100 µl freshly made elution buffer to final pellet and rotate 15 min at room temperature, sediment agarose as before, and transfer supernatant (containing ChIP DNA) to a new tube.
9. Add an additional 100 µl elution buffer to pellet, repeat elution, and combine two eluates.
10. Add 8 µl 5 M NaCl per 200 µl eluate and incubate at 65 °C overnight.
11. On the next day, add 8 µl 1 M Tris–HCl, pH 7.5, 4 µl 0.5 M EDTA, and 1 µl 10 mg/ml proteinase K to each tube, incubate 1 h at 45 °C.
12. Purify DNA using Qiagen PCR purification kit, and elute in 50 µl EB. To validate ChIP, perform qRT-PCR to assess the enrichment of ChIP DNA relative to input DNA within positive and negative control genomic regions, and calculate enrichment ratio $\frac{C+/I+}{C-/I-}$. *C*+ and *I*+ are ChIP and input C_T values at the positive control, whereas *C*– and *I*– are ChIP and input C_T values at negative control. Enrichment ratios greater than 10 are indicative of successfully ChIP.

3.2.3 Preparation of ChIP-seq Library and High-Throughput Sequencing

1. To prepare a ChIP-seq library, start with 40 µl of ChIP DNA and input DNA from Subheading 3.2.2, and use NEBNext ChIP-seq Library Prep Reagent Set to perform end repair, dA-tailing of end-repaired DNA, adaptor ligation, and PCR enrichment of DNA library/multiplexing. AMPure XP beads are used to size select and cleanup DNA between each step.
2. Assess library quality as described in Subheading 3.1.3 and subject samples to sequencing.

4 Notes

1. Either FACS using antibodies against cell surface antigens or intracellular hormones can be used to isolate high-purity beta cells and alpha cells. For delta cell purification, however, due to the lack of specific cell surface marker, FACS using from fixed and permeabilized islet cells and an antibody against Somatostatin is currently the most reliable means.
2. Resuspend in 1 ml 0.05% trypsin–EDTA first, and then add additional 2 ml 0.05% trypsin–EDTA; mix again.
3. To help content pass through the strainer, gently press a p1000 tip against the strainer membrane in a swirling motion.
4. Can pause at this point for a few hours.
5. As secondary antibodies are conjugated with light-sensitive fluorophores, for all incubation and washing steps onward, wrap tubes with aluminum foil to prevent quenching of fluorescent signals.
6. Human islet cells have notoriously high autofluorescence. It is important to have unstained cells from each donor as a negative control to ensure precise gating parameters.
7. If sorted cells are intended for ChIP, cross-link them in 500 μl "fix buffer" containing 1% formaldehyde for exactly 10 min at room temperature on a rotator. Stop cross-linking by adding 20 μl 2.5 M glycine (190 mg in 1 ml H_2O) to every 500 μl fix buffer, and rotate at room temperature for 5 min. Prolonged fixation impedes downstream chromatin shearing.
8. After fixation/permeabilization, cells are more buoyant and resistant to centrifugation. Subsequent centrifugation steps occur at this higher g-force for shorter duration to reduce time during which RNA degradation may occur. During wash steps, take care when removing supernatant so as to not aspirate a thin layer of cell pellet on the side of the tube.
9. Low retention tubes (1.5 ml) can be used to collect cells after FACS to minimize cell loss.
10. Flow-through can be snap frozen and stored at −80 °C for future use.
11. If low cell numbers are collected ($<10^5$ cells), elute RNA still bound to the column for the second time with 60 μl of nuclease-free water. In addition, RNA can be vacuum concentrated if needed.
12. The reaction volume can be scaled up if RNA concentration is low.
13. The Ovation RNA-seq System V2 offers two advantages to circumvent low sample quality and abundance. (a) RNA amplification is initiated at the 3′ end as well as randomly throughout

the whole transcriptome in the sample, so low-integrity RNA can also be successfully amplified. (b) The final cDNA amplification step usually produces enough material for both qRT-PCR assays and RNA-seq from one sample.

14. If peaks around 80 bp (primers) or 128 bp (adaptor dimer) are present in the Bioanalyzer traces, bring up the sample volume to 50 μl exactly with nuclease-free water and repeat the AMPure XP bead clean up steps.

Acknowledgments

We are grateful for technical support from the Next-Generation Sequencing Core (https://ngsc.med.upenn.edu/) and the Flow Cytometry and Cell Sorting Facility at the University of Pennsylvania. We also thank Dr. Sinisa Hrvatin, former graduate student from Dr. Doug Melton's lab, for sharing his protocol of intracellular marker sorting. We appreciate the valuable discussion from the Kaestner lab members: Dr. Amanda Ackermann, Julia Kieckhaefer, Diana Bernstein, and Dr. Julia Wang. This work was supported by NIH grant UC4DK104119 and ADA grant 7-12-MN-37.

References

1. Bramswig NC, Kaestner KH (2012) Epigenetics and diabetes treatment: an unrealized promise? Trends Endocrinol Metab 23(6):286–291. doi:10.1016/j.tem.2012.02.002
2. Dawson MA, Kouzarides T (2012) Cancer epigenetics: from mechanism to therapy. Cell 150(1):12–27. doi:10.1016/j.cell.2012.06.013
3. Pinney SE, Simmons RA (2010) Epigenetic mechanisms in the development of type 2 diabetes. Trends Endocrinol Metab 21(4):223–229. doi:10.1016/j.tem.2009.10.002
4. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12(1):7–18. doi:10.1038/nrg2905
5. Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH (2015) TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest 125(5):1998–2006. doi:10.1172/JCI77321
6. Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, Kaestner KH (2013) Epigenomic plasticity enables human pancreatic alpha to beta cell reprogramming. J Clin Invest 123(3):1275–1284. doi:10.1172/JCI66514
7. Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, Harlan DM, Powers AC (2005) Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 53(9):1087–1097. doi:10.1369/jhc.5C6684.2005
8. Dorrell C, Abraham SL, Lanxon-Cookson KM, Canaday PS, Streeter PR, Grompe M (2008) Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. Stem Cell Res 1(3):183–194. doi:10.1016/j.scr.2008.04.001
9. Hrvatin S, Deng F, O'Donnell CW, Gifford DK, Melton DA (2014) MARIS: method for analyzing RNA following intracellular sorting. PLoS One 9(3), e89459. doi:10.1371/journal.pone.0089459

Chapter 9

eIF3 Regulation of Protein Synthesis, Tumorigenesis, and Therapeutic Response

Ji-Ye Yin, Zizheng Dong, and Jian-Ting Zhang

Abstract

Translation initiation is the rate-limiting step of protein synthesis and highly regulated. Eukaryotic initiation factor 3 (eIF3) is the largest and most complex initiation factor consisting of 13 putative subunits. A growing number of studies suggest that eIF3 and its subunits may represent a new group of proto-oncogenes and associates with prognosis. They regulate translation of a subset of mRNAs involved in many cellular processes including proliferation, apoptosis, DNA repair, and cell cycle. Therefore, unveiling the mechanisms of eIF3 action in tumorigenesis may help identify attractive targets for cancer therapy. Here, we describe a series of methods used in the study of eIF3 function in regulating protein synthesis, tumorigenesis, and cellular response to therapeutic treatments.

Key words Eukaryotic initiation factor 3 (eIF3), Translational control, Protein synthesis, Tumorigenesis, Therapeutic response

1 Introduction

Gene expression is regulated primarily at levels of transcription and translation. Translation is a process of transferring genetic information from mRNA to protein, and, thus, its deregulation results in abnormal gene expression. Aberrant protein synthesis and dysregulation of mRNA translation have been associated with diseased state such as cancer [1–5]. In eukaryotes, mRNA translation is a complicated process consisting of three major steps: initiation, elongation, and termination [6]. Translation initiation is the rate-limiting step of protein synthesis and therefore highly regulated. Eukaryotic initiation factors (eIFs) are proteins that play important roles in the initiation step.

Among all human eIFs, eIF3 is the largest and most complex one consisting of 13 putative subunits, which are named as eIF3a to eIF3m. A growing number of studies suggest that eIF3 subunits may associate with tumorigenesis and therapeutic response through

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_9,

regulating protein synthesis [3, 7, 8]. Here, we describe a series of techniques and methods used in the study of eIF3 in regulating protein synthesis, tumorigenesis, and therapeutic response.

2 Materials

2.1 Reagents

1. EDTA.
2. Goat and fetal bovine serum.
3. H_2O_2.
4. Phosphate-Buffered Saline (PBS).
5. Bovine serum albumin (BSA).
6. Propidium iodide (25 μg/ml).
7. AMV reverse transcriptase.
8. Reverse Transcription System Kit (Promega, Madison, WI, USA).
9. Fluorescent SYBR Green dyes.
10. Culture media supplemented with 10% serum.
11. Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA).
12. Dimethyl sulfoxide (DMSO).
13. Annexin V-FITC Apoptosis Detection Kit I (BD Pharmingen, San Jose, CA, USA).
14. DAPI or Hoechst-33342 dye.
15. [^{35}S]methionine.
16. Trichloroacetic acid (TCA).
17. Cycloheximide (CHX).
18. RNeasy Mini Kit (Qiagen, Hilden, Germany).
19. T7 or SP6 RNA polymerase.
20. rNTP.
21. Dithiothreitol (DTT).
22. RNasin.
23. m^{7}GpppG.
24. DNase I.
25. Rabbit reticulocyte lysate (RRL) system (Promega, Madison, WI, USA).
26. Opti-MEM I (Invitrogen, Carlsbad, CA, USA).
27. Lipofectin (Invitrogen, Carlsbad, CA, USA).
28. Passive Lysis Buffer (Promega, Madison, WI, USA).
29. Luciferase Reporter Assay Kit (Promega, Madison, WI, USA).
30. α[^{32}P]-UTP.
31. RNase mixture (4 μg/μl RNase A and 5 U/μl RNase T1).

32. Biotin-11-CTP.
33. Streptavidin MagneSphere Paramagnetic Particles (Promega, Madison, WI, USA).
34. Phenol/chloroform/isoamyl alcohol.
35. Ammonium acetate.
36. Lithium chloride.
37. Glycogen.
38. 100% ethanol.
39. RNase-free water.

2.2 Buffers and Solutions

1. Lysis buffer 1 [50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1% NP-40, 1 mM sodium orthovanadate, 1 mM EDTA, 1 mM sodium fluoride, 100 mg/ml phenylmethylsulfonyl fluoride (PMSF), 100 mg/ml DTT].
2. 1× binding buffer 1 [10 mM Hepes/NaOH (pH 7.4), 140 mM NaCl, 2.5 mM $CaCl_2$].
3. Polysome extraction buffer [10 mM Tris–HCl (pH 7.4), 150 mM NaCl, 5 mM $MgCl_2$, 1% Triton X-100, 40 mM DTT, 0.1 mg/ml CHX, and 1 mg/ml heparin].
4. Binding buffer 2 [20 mM Hepes (pH 7.9), 100 mM KCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, and 5 mM $MgCl_2$].
5. Wash buffer 1 (10 mM (Hepes), 20 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, and 1 mM PMSF).
6. Lysis buffer 2 (10 mM Hepes, pH 7.0, 100 mM KCl, 5 mM $MgCl_2$, 0.5% NP-40, 1 mM DTT, 100 units/ml RNase inhibitor, protease inhibitor cocktail).
7. Wash buffer 2 (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40).
8. Immunoprecipitation buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40, 20 mM EDTA, pH 8.0, 1 mM DTT, 100 units/ml RNase inhibitor).
9. Proteinase K solution (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40, 1% SDS, 1.2 mg/ml proteinase K).
10. Crystal violet staining solution (0.5% crystal violet in 20% methanol).
11. 0.6% and 0.3% agar solution in cell culture medium.
12. 10% formalin buffer.
13. 5 mg/ml thiazolyl blue tetrazolium bromide solution in 1× PBS.

2.3 Plasmids and siRNAs

1. pCβA vector control.
2. pCβA-eIF3 subunit.
3. eIF3 subunit siRNAs.
4. Scrambled control siRNAs.

3 Methods

3.1 Detection of EIF3 Expression and Localization

Expression of various eIF3 subunits has been found to be altered in various human cancers [7], and it has also been found that eIF3a may localize in different subcellular compartments [9, 10]. Thus, to investigate the expression of eIF3 subunits and their subcellular localization, it is necessary to detect these proteins effectively in either cell line models or in clinical samples. For studies involving manipulation of eIF3 expression using ectopic overexpression or RNA interference knockdowns, it is also necessary to effectively detect these proteins using different methods.

3.1.1 Western Blot

1. Cells are lysed in lysis buffer 1 [50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1% NP-40, 1 mM sodium orthovanadate, 1 mM EDTA, 1 mM sodium fluoride, 100 mg/ml phenylmethylsulfonyl fluoride (PMSF), 100 mg/ml dithiothreitol (DTT)] at 4 °C for 30 min.
2. Protein concentration of the lysate is determined using the Bradford method.
3. Then, protein samples are separated by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to polyvinylidene difluoride (PVDF) membranes.
4. The membranes are then blocked with 5% nonfat milk and incubated with eIF3 subunit antibodies overnight followed by washing and incubation with horseradish peroxidase-conjugated secondary antibodies.
5. The reaction is detected using ECL reagents, and the signals are captured by X-ray films or by imaging systems.

3.1.2 Immunohistochemistry (IHC)

1. To conduct IHC staining, serial 4-mm thick sections are cut from tissue blocks and mounted on slides. The slide containing maximum amount of tumors is selected for each case, and one representative slide from each case is used.
2. Sample sections on slides are first baked at 60 °C for 30 min followed by incubation in xylene for 2×10 min and rehydration through graded ethanol to distilled water.
3. Antigen retrieval is done by heating samples in 1 mmol/L EDTA (pH 8.0) for 20 min.

4. Nonspecific staining is blocked by 10% goat serum in PBS buffer for 20 min at room temperature, and endogenous peroxidase activity is quenched by incubation in 3% H_2O_2 for 10 min.
5. Slides are then incubated with eIF3 subunit antibody/antibodies or PBS control at 4 °C overnight followed by incubation with biotinylated secondary antibody and peroxidase-conjugated streptavidin.
6. The staining is then visualized by using 3,3′-diaminobenzidine tetrahydrochloride substrate, and all samples are counterstained with hematoxylin and eosin (H&E) before viewing with an inverted microscope (*see* **Note 1**).

3.1.3 Immunofluorescence (IF)

1. For immunofluorescence staining, cells are seeded on a glass coverslip in a six-well plate and allowed to grow until near confluence.
2. The cells on cover glass are washed three times with ice-cold phosphate-buffered saline and fixed with acetone/methanol (1:1) at room temperature for 10 min and incubated at 4 °C for 30 min with blocking solution (1% BSA in PBS).
3. The cells are then probed with eIF3 antibodies for 1 h at 4 °C followed by incubation with secondary antibody conjugated with fluorescein isothiocyanate (FITC) at 4 °C for another 1 h.
4. After washing three times with blocking solution, cell nuclei are counterstained with propidium iodide (25 μg/ml) for 10 min (*see* **Note 2**).
5. The coverslips are then mounted on the slides and viewed with a confocal microscope.

3.1.4 Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

1. Firstly, total RNA is isolated using RNeasy Mini Kit (Qiagen, Hilden, Germany).
2. 1 μg of the total RNAs is reverse-transcribed with 0.5 μg random primers and 15 unit AMV reverse transcriptase in a total volume of 20 μl at 42 °C for 1 h using Reverse Transcription System Kit (Promega, Madison, WI, USA).
3. The real-time PCR is then carried out in a real-time PCR system with fluorescent SYBR Green dyes.
4. The cycle threshold value (C_t) is defined as the PCR cycle number at which the reporter fluorescence crosses the threshold. The C_t of each product is determined and normalized against that of the internal control (β-actin or GAPDH).

3.2 EIF3-Induced Tumorigenic Assays

A large number of studies showed that some eIF3 subunits may play important roles in tumorigenesis, including eIF3a, eIF3b, eIF3c, eIF3h, and eIF3i, using ectopic overexpression and siRNA knockdown [11–22] and assays detecting the effect of eIF3 on cell growth, apoptosis, malignant transformation, and tumor formation.

3.2.1 Ectopic Overexpression and Knockdown of eIF3

1. Cells are seeded in a 6-well plate and maintained in culture media supplemented with 10% serum (*see* **Note 3**).
2. It is followed by transfection with 4 μg pCβA vector control and eIF3 subunit-expressing constructs for overexpression or 50 nmol/L eIF3 subunit and scrambled control siRNAs for knockdown using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA).
3. Following incubation for 48 h post-transfection, cells can be harvested for further analysis.
4. To establish stable clones with eIF3 subunit overexpression, the transiently transfected cells are collected 24 h post-transfection and replated in 100-mm dishes followed by selection with 0.6–1.0 mg/mL G418 for 2 weeks.
5. The G418-resistant clones are further propagated for testing eIF3 expression using real-time PCR and Western blot analyses as described above. The positive stable clones are then maintained in the presence of 0.2 mg/mL G418 for further analysis.

3.2.2 Colony Formation and Anchorage-Independent Growth Assay

1. Colony formation assay is performed by seeding 100 cells per well in six-well plates and cultured for 10–14 days with medium changed every 2–3 days.
2. At the end of the assay, cell colonies are stained with crystal violet (0.5% crystal violet in 20% methanol) for 20 min and washed thoroughly with water. The visible colonies are counted manually.
3. For anchorage-independent growth assay, cells are suspended at a density of 2.5×10^3 cells/ml in 0.3% agar solution in cell culture medium.
4. 1 ml of this suspension is overlaid on top of a 0.6% agar layer made in cell culture medium in a six-well plate.
5. Cells are then cultured for 14–25 days with fresh medium changed every 2–3 days.
6. At the end of assays, the cells are stained as described above and visible colonies are counted manually.

3.2.3 Xenograft Tumor Formation Assay

1. Xenograft tumor formation in immune deficient mice (*see* **Note 4**) is an essential assay to investigate the tumorigenesis function of eIF3 genes.
2. Approximately 1×10^7 cells are injected subcutaneously into 7-week-old nonobese diabetic/severe combined immunodeficient (NOD/SCID) or nude mice (one injection/mouse).
3. Tumor growth is measured by a caliper twice a week for a total of 4–9 weeks. The tumor volume is calculated from two perpendicular diameters using the formula: $\text{volume} = (\text{length}/2) \times (\text{width}^2)$.
4. To confirm the xenograft tumor pathology, tumor tissues are removed, measured, and fixed in 10% formalin buffer.

5. After staining with H&E, samples are subjected to standard histology and pathology analyses.

3.3 Therapeutic Response

In addition to the potential role of eIF3 in tumorigenesis, eIF3 has also been found to associate with prognosis [14, 23–26] possibly by regulating cellular response to anticancer drugs [14, 27]. Thus, assays to analyze cellular response to different anticancer drugs are important for studying the role of eIF3 in cell survival against anticancer drugs and in drug-induced apoptosis.

3.3.1 Survival Methyl Thiazolyl Tetrazolium (MTT) Assay

1. MTT assay is the easiest and one of the most commonly used approaches for studying cell survival.
2. Cells are seeded in 96-well plates at a density of 2000 cells/well and cultured for 24 h before treating the cells with anticancer drugs at different concentrations for 3 days.
3. The culture medium is then removed, and thiazolyl blue tetrazolium bromide is added to a final concentration of 0.5 mg/ml followed by incubation for 4 h at 37 °C.
4. The formazan is then solubilized by adding 150 μl/well dimethyl sulfoxide (DMSO) and OD_{570nm} is measured in a microplate reader.
5. The half of maximal inhibitory concentration (IC_{50}) values can be obtained from the dose–response curves.
6. Alternatively, colony formation assay as described above can be used to investigate survival following drug treatments.

3.3.2 Apoptosis Assay

Apoptosis can be detected using several assays, and in the following section, we describe two assays that have been used to study eIF3. Annexin V staining is a most commonly used method to quantitatively measure apoptotic cell populations:

1. Firstly, cells are seeded in 6-well plate and allowed to attach overnight.
2. Then, the medium is replaced with fresh one containing anticancer drugs and the cells are cultured for 24 h.
3. The cells are then washed twice with ice-cold phosphate-buffered saline and resuspended in 1× binding buffer 1 [10 mM Hepes/NaOH (pH 7.4), 140 mM NaCl, 2.5 mM $CaCl_2$] at 1×10^6 cells/ml and incubated with FITC-conjugated Annexin V antibody and propidium iodide counterstain for DNA using Annexin V-FITC Apoptosis Detection Kit I (BD Pharmingen, San Jose, California, USA).
4. About 10,000 cells in each sample are analyzed using a BD FACSCalibur flow cytometer. Cell apoptosis profiles are analyzed using BD CellQuest^{Pro} software.

5. Another assay is based on the staining of disintegrated nuclei by using DAPI or Hoechst-33342.
6. Briefly, cells following anticancer drug treatment are harvested and stained with 1 μg/ml DAPI or 1 μg/ml Hoechst-33342 dye.
7. The stained cells are then mounted onto a polylysine-coated slide by centrifugation and examined under a fluorescent microscope.
8. A total of 300–400 nuclei from five randomly chosen fields are examined, and the nuclei displaying distinctive apoptosis-associated morphological changes are scored. Apoptosis can be quantified as a fraction of the total number of nuclei examined.

3.4 Protein Synthesis Assays

One mechanism of eIF3 regulation of cell proliferation and response to insults is that they differentially regulate translation of a subset of specific cancer and survival-related mRNAs [13, 28, 29]. Several assays can be used to detect translational regulation.

3.4.1 Metabolic Labeling

1. [^{35}S]methionine labeling in live cells is widely used to measure total or specific protein synthesis.
2. Briefly, 1×10^5 cells per well are seeded in a 24-well plate in triplicates and incubated for 24 h followed by washing three times with serum-free and methionine-free medium.
3. The cells are then incubated with methionine-free media supplemented with 20 μCi/ml [^{35}S]methionine for 30 min, washed for three times, and harvested.
4. Next, cells are lysed followed by separation on SDS-PAGE and autoradiography analysis for determination of global protein synthesis or immunoprecipitation of specific proteins of interest and then separation on SDS-PAGE and autoradiography analysis of specific proteins.
5. Alternatively, cell lysates are precipitated with 10% TCA. The acid-insoluble material is collected on a filter by rapid filtration, and the radioactivity is determined by scintillation counting to quantify global protein synthesis.

3.4.2 Polysome Profiling Analysis

1. Polysomes are defined as mRNA bound with multiple ribosomes during translation; it represents actively translated mRNAs [30]. Thus, analyzing polysomes using high-throughput methods enables profiling of actively translated mRNAs.
2. Approximately 5×10^7 cells are incubated at 37 °C prior to the experiment. Cycloheximide (CHX) is firstly added into the medium at a final concentration of 0.1 mg/ml for 10 min at 37 °C.

3. The medium is then removed and cells are washed twice with cold PBS containing 0.1 mg/ml CHX followed by harvesting cells.
4. The cells are lysed using 500 µl polysome extraction buffer [10 mM Tris–HCl (pH 7.4), 150 mM NaCl, 5 mM $MgCl_2$, 1 % Triton X-100, 40 mM DTT, 0.1 mg/ml CHX, and 1 mg/ml heparin].
5. Extracts are incubated on ice for 15 min with occasional vortexing followed by centrifugation to remove nuclei and debris. 500 µl supernatant is recovered and layered onto 10 ml, 10–50 % linear sucrose gradients in extraction buffer lacking Triton X-100.
6. The gradients are centrifuged at 38,000 × *g* for 3 h at 4 °C.
7. Then polysome profiles are monitored by collecting fractions and concomitant measurement of the absorbance at 254 nm. Polysomal RNAs are extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany) and subjected to further analysis (*see* **Note 5**).

3.4.3 *In Vitro Transcription and Translation*

1. In vitro translation in rabbit reticulocyte lysate can be programed using cRNAs, which can also be generated using in vitro transcription [31–34].
2. To perform in vitro transcription, cDNA templates are linearized by digestion with restriction endonuclease, and cRNA transcripts carrying both 5′-cap and 3′-poly(A) tail are synthesized by incubating 15 units T7 or SP6 RNA polymerase in the presence of 5 µg linearized DNA, 0.5 mM rNTP, 10 mM DTT, 60 units RNasin, and 30 mM m^7GpppG in a final volume of 50 µl at 37 °C for 1.5 h.
3. After digestion with 3 units DNase I at 37 °C for 15 min, the in vitro cRNA transcripts are then purified using RNeasy Mini Kit (Qiagen, Hilden, Germany).
4. For in vitro translation, about 50 ng of in vitro capped cRNA transcripts are used to program cell-free translation in rabbit reticulocyte lysate (RRL) system (Promega, Madison, WI, USA) in a final volume of 10 µl containing 3.5 µl RRL in the presence of [^{35}S]methionine and purified eIF3a proteins followed by SDS-PAGE separation and autoradiography analyses.

3.4.4 *RNA-Based Luciferase Reporter Assay*

1. RNA-Based Luciferase Reporter Assay can quantitatively measure translational regulation in a convenient way and to study the role of UTRs in translational control.
2. For this assay, in vitro transcripts with specific UTRs of interests are advised to be used in place of DNA constructs encoding the transcripts.
3. Briefly, 2×10^5 cells/well are seeded into 6-well plates on the day before transfection. Cells are then washed once with Opti-

MEM I (Invitrogen, Carlsbad, CA, USA) with reduced serum medium and incubated with a mixture containing 12.5 μg Lipofectin (Invitrogen, Carlsbad, CA, USA), 1 μg transcripts, and 1 ml Opti-MEM I medium.

4. At 8 h after transfection, cells are lysed in 500 μl 1× Passive Lysis Buffer (Promega, Madison, WI, USA). The luciferase activities can be determined using Luciferase Reporter Assay Kit (Promega, Madison, WI, USA).

3.5 EIF3-mRNA Binding Assays

eIF3 proteins have been reported to regulate translation of a subset of specific cancer-related mRNAs possibly by directly binding to these mRNAs [11, 29, 31]. Therefore, assays to assess eIF3–nucleic acid interactions are important for understanding the molecular mechanism of eIF3 function in translational regulation.

3.5.1 UV Cross-Linking

1. UV cross-linking is a powerful method to detect RNA-binding proteins [11, 31]. UV irradiation can trigger the formation of covalent bond between RNA and its binding proteins.
2. RNA probes are firstly generated using in vitro transcription as described above in the presence of 70 μCi α[^{32}P]-UTP and purified using RNeasy Mini Kit (Qiagen, Hilden, Germany).
3. The RNA probe is then diluted to 1×10^5 cpm/μl and mixed with 20 μg purified eIF3 proteins or total cell lysate and 30 μg *E. coli* tRNA in binding buffer 2 [20 mM Hepes (pH 7.9), 100 mM KCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, and 5 mM $MgCl_2$] followed by incubation at RT for 30 min and irradiation by UV (254 nm, 5.4 J/cm^2).
4. Following digestion using 10 μl RNase mixture (4 μg/μl RNase A and 5 U/μl RNase T1) at 37 °C for 30 min, the mixture can be separated by SDS-PAGE for autoradiography if purified protein is used.
5. In the case of cell lysate, the reaction mixture is subjected to immunoprecipitation of the eIF3 protein of interest followed by SDS-PAGE separation and autoradiography.

3.5.2 RNA Electrophoretic Mobility Shift Assay (EMSA)

1. EMSA is a standard affinity electrophoresis method to detect protein–nucleic acid (DNA or RNA) interactions [11, 35, 36].
2. For RNA EMSA, RNA probes are firstly generated using in vitro transcription in the presence of α[^{32}P]-UTP and purified by using RNeasy Mini Kit as described above. Then 1–5 μg purified proteins or total cell lysate is mixed with 200 μg/mL of yeast tRNA, in binding buffer [10 mM HEPES (pH 7.9), 50 mM KCl, 10% glycerol, 0.2 mg/ml BSA, 1 mM DTT, and 0.2 mM PMSF], and 4×10^4 cpm α[^{32}P]-labeled RNA probes.
3. The mixture is incubated for 30 min at room temperature and unbound probes are digested by 100 units RNase T1 for

15 min at 30 °C. The reaction mixtures are then separated on non-denaturing PAGE. The signal is detected by autoradiography (*see* **Note 6**).

3.5.3 Pulldown Assay Using Biotinylated RNA Probe

1. Pulldown assays using biotinylated RNA probes have an advantage of avoiding the use of radio isotopes [11, 31].
2. The biotinylated RNA probes are generated using in vitro transcription as described above but in the presence of 0.625 mM Biotin-11-CTP and purified as described above.
3. Biotinylated RNA probe is incubated with 20 μg purified eIF3 proteins or total cell lysate and 30 μg *E. coli* tRNA in binding buffer 2 [20 mM Hepes (pH 7.9), 100 mM KCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, and 5 mM $MgCl_2$] at RT for 1 h and followed by UV irradiation (254 nm, 5.4 J/cm^2).
4. The RNA–protein complexes are then digested by 10 μl RNase mix (4 μg/μl RNase A and 5 U/μl RNase T1) at 37 °C for 30 min and isolated using 0.5 mg/ml Streptavidin MagneSphere Paramagnetic Particles (Promega, Madison, WI, USA) at room temperature for 30 min, followed by washing for three times with wash buffer 1 (10 mM *Hepes*, 20 mM KCl, 1 mM $MgCl_2$, 1 mM dithiothreitol, and 1 mM PMSF).
5. The pulldown materials are then separated by SDS-PAGE and analyzed using Western blot probed by antibodies specific to the eIF3 proteins of interest.

3.5.4 RNA Immunoprecipitation

1. The above methods are directed for studies using in vitro transcribed RNA probes. To investigate interaction between endogenous proteins and RNAs, RNA immunoprecipitation can be used.
2. Firstly, cells are lysed in lysis buffer 2 (10 mM Hepes, pH 7.0, 100 mM KCl, 5 mM $MgCl_2$, 0.5% NP-40, 1 mM DTT, 100 units/ml RNase inhibitor, protease inhibitor cocktail) by pipetting up and down 7–10 times followed by incubation on ice for 5 min and then freezing at −80 °C overnight.
3. Next day, protein A-conjugated magnetic beads are washed with wash buffer 2 (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40) for one time and then resuspended in 100 μl wash buffer containing 5 μg eIF3 antibody followed by incubation for 30 min at room temperature, washing for three times, and resuspended in 900 μl immunoprecipitation buffer (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40, 20 mM EDTA, pH 8.0, 1 mM DTT, 100 units/ml RNase inhibitor).
4. The cell lysate is thawed quickly and cleared by centrifugation at 15,000 × *g* for 10 min at 4 °C. Aliquot 10 μl of the cell lysate

and mark it as input and store it at −80 °C. Mix 100 μl cell lysate to the bead suspension above and incubate at 4 °C for 4–6 h or overnight with agitation. At the end of incubation, the beads are collected and washed with the wash buffer for 6 times. Finally, the beads are resuspended in 150 μl proteinase K solution (50 mM Tris–HCl, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40, 1% SDS, 1.2 mg/ml proteinase K), while the cell lysate input is thawed and mixed with 140 μl proteinase K solution, followed by incubation at 55 °C for 45 min and mixing with 250 μl wash buffer to make the final volume to 400 μl.

5. The mixtures are then extracted with phenol/chloroform/isoamyl alcohol. The RNA is precipitated by adding 1/6 volume 5 M ammonium acetate, 1/20 volume 7.5 M lithium chloride, 1/60 volume 5 mg/ml glycogen, and 2.5 volume ethanol and storing at −80 °C overnight. The RNAs are recovered and dissolved in 20 μl RNase-free water.
6. Take 6 RNA μL samples to do reverse transcription to synthesize cDNA and analyze by quantitative PCR.

4 Notes

1. Quantification of IHC staining is difficult but important to be performed. All IHC staining of tissue sections should be evaluated independently by at least two pathologists. Microscopic fields with the highest degree of immunoreactivity should be chosen for analysis, and at least 1000 cells need to be analyzed in each case. The score of cells exhibiting staining in each case is evaluated semiquantitatively. A numeric intensity score is set from 1 to 4 (1 for no, 2 for weak, 3 for moderate, and 4 for strong staining). The fraction score (0–100%) is defined by the percentage of positive tumor cells per slide. Total score range of 0–400 is obtained by multiplying the intensity score and the fraction score. The scores can be used to conduct statistical analysis as both continuous and binary variable by defining high and low expression levels.
2. Here, we described the use of FITC-conjugated secondary antibody and propidium iodide for staining. Other fluorescein (e.g., rhodamine)-labeled antibodies and nuclei staining dye (e.g., DAPI) can also be used.
3. Ectopic and knockdown expression should be performed in eIF3 low and high expression cells, respectively. For example, immortalized non-cancer cell lines such as NIH3T3, RIE, IEC, and IMR-90 cells with low eIF3a expression can be used to establish eIF3a overexpression cell lines, while cancer cell lines such as H1299, A549, HeLa, and MCF7 cells with high

endogenous eIF3a expression can be used to knockdown eIF3a expression [11, 13, 14, 27, 28, 37].

4. It is noteworthy that in vivo studies using animal models need to be approved by institutional animal care and use committee before initiation of animal study.
5. Since polysome profiling enables analysis of the translational level of a large number of mRNAs, high-throughput methods should be used to detect these mRNAs. Previously, microarray is the major method to conduct the high-throughput analysis [6, 38]. With the rapid development of next-generation sequencing, RNA-seq is replacing microarray and becoming a powerful method to analyze polysome profiling. It is, thus, recommended.
6. For supershift and competition, 2 μL of specific antibodies against target eIF3 proteins or 100-fold cold probe is added to the reaction mixture and incubate for 30 min before adding α[^{32}P]-labeled probe.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China Grants 81573463, Hunan Provincial Natural Science Foundation of China Grant 2015JJ1024, and National Institutes of Health Grant R01 CA140582.

References

1. Hershey JWB, Miyamoto S (2000) Translational control and cancer. In: Sonenberg N, Hershey JWB, Mathews MB (eds) Translational control of gene expression. Cold Spring Harbor Laboratories Press, New York, pp 637–654
2. Silvera D, Formenti SC, Schneider RJ (2010) Translational control in cancer. Nat Rev Cancer 10(4):254–266. doi:10.1038/nrc2824
3. Yin JY, Dong Z, Liu ZQ, Zhang JT (2011) Translational control gone awry: a new mechanism of tumorigenesis and novel targets of cancer treatments. Biosci Rep 31(1):1–15. doi:10.1042/BSR20100077, BSR20100077 [pii]
4. Ruggero D (2013) Translational control in cancer etiology. Cold Spring Harb Perspect Biol 5(2). doi:10.1101/cshperspect.a012336
5. Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J, Topisirovic I (2015) Targeting the translation machinery in cancer. Nat Rev Drug Discov 14(4):261–278. doi:10.1038/nrd4505
6. Sonenberg N, Hershey JWB, Mathews M (2000) Translational control of gene expression, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
7. Hershey JW (2015) The role of eIF3 and its individual subunits in cancer. Biochim Biophys Acta 1849(7):792–800. doi:10.1016/j.bbagrm.2014.10.005
8. Dong Z, Zhang JT (2006) Initiation factor eIF3 and regulation of mRNA translation, cell growth, and cancer. Crit Rev Oncol Hematol 59(3):169–180
9. Shen J, Yin JY, Li XP, Liu ZQ, Wang Y, Chen J, Qu J, Xu XJ, McLeod HL, He YJ, Xia K, Jia YW, Zhou HH (2014) The prognostic value of altered eIF3a and its association with p27 in non-small cell lung cancers. PLoS One 9(4), e96008. doi:10.1371/journal.pone.0096008
10. Pincheira R, Chen Q, Huang Z, Zhang JT (2001) Two subcellular localizations of eIF3 p170 and its interaction with membrane-bound microfilaments: implications for alternative functions of p170. Eur J Cell Biol 80(6):410–418

11. Qi J, Dong Z, Liu J, Zhang JT (2014) EIF3i promotes colon oncogenesis by regulating COX-2 protein synthesis and beta-catenin activation. Oncogene 33(32):4156–4163. doi:10.1038/onc.2013.397
12. Liu Z, Dong Z, Yang Z, Chen Q, Pan Y, Yang Y, Cui P, Zhang X, Zhang JT (2007) Role of eIF3a (eIF3 p170) in intestinal cell differentiation and its association with early development. Differentiation 75(7):652–661. doi:10.1111/j.1432-0436.2007.00165.x, DIF165 [pii]
13. Dong Z, Liu LH, Han B, Pincheira R, Zhang JT (2004) Role of eIF3 p170 in controlling synthesis of ribonucleotide reductase M2 and cell growth. Oncogene 23(21):3790–3801
14. Yin JY, Shen J, Dong ZZ, Huang Q, Zhong MZ, Feng DY, Zhou HH, Zhang JT, Liu ZQ (2011) Effect of eIF3a on response of lung cancer patients to platinum-based chemotherapy by regulating DNA repair. Clin Cancer Res 17(13):4600–4609. doi:10.1158/1078-0432.CCR-10-2591, 1078-0432.CCR-10-2591 [pii]
15. Zhang L, Pan X, Hershey JW (2007) Individual overexpression of five subunits of human translation initiation factor eIF3 promotes malignant transformation of immortal fibroblast cells. J Biol Chem 282(8):5790–5800. doi:10.1074/jbc.M606284200, M606284200 [pii]
16. Wang H, Ru Y, Sanchez-Carbayo M, Wang X, Kieft JS, Theodorescu D (2013) Translation initiation factor eIF3b expression in human cancer and its role in tumor growth and lung colonization. Clin Cancer Res 19(11):2850–2860. doi:10.1158/1078-0432.CCR-12-3084
17. Rasmussen SB, Kordon E, Callahan R, Smith GH (2001) Evidence for the transforming activity of a truncated Int6 gene, in vitro. Oncogene 20(38):5291–5301. doi:10.1038/sj.onc.1204624
18. Mayeur GL, Hershey JW (2002) Malignant transformation by the eukaryotic translation initiation factor 3 subunit p48 (eIF3e). FEBS Lett 514(1):49–54
19. Shi J, Kahle A, Hershey JW, Honchak BM, Warneke JA, Leong SP, Nelson MA (2006) Decreased expression of eukaryotic initiation factor 3f deregulates translation and apoptosis in tumor cells. Oncogene 25(35):4923–4936. doi:10.1038/sj.onc.1209495
20. Zhang L, Smit-McBride Z, Pan X, Rheinhardt J, Hershey JW (2008) An oncogenic role for the phosphorylated h-subunit of human translation initiation factor eIF3. J Biol Chem 283(35):24047–24060. doi:10.1074/jbc.M800956200
21. Wang YW, Lin KT, Chen SC, Gu DL, Chen CF, Tu PH, Jou YS (2013) Overexpressed-eIF3I interacted and activated oncogenic Akt1 is a theranostic target in human hepatocellular carcinoma. Hepatology 58(1):239–250. doi:10.1002/hep.26352
22. Ahlemann M, Zeidler R, Lang S, Mack B, Munz M, Gires O (2006) Carcinoma-associated eIF3i overexpression facilitates mTOR-dependent growth transformation. Mol Carcinog 45(12):957–967. doi:10.1002/mc.20269
23. Chen G, Burger MM (1999) p150 expression and its prognostic value in squamous-cell carcinoma of the esophagus. Int J Cancer 84(2):95–100
24. Dellas A, Torhorst J, Bachmann F, Banziger R, Schultheiss E, Burger MM (1998) Expression of p150 in cervical neoplasia and its potential value in predicting survival. Cancer 83(7):1376–1383
25. Chen G, Burger MM (2004) p150 overexpression in gastric carcinoma: the association with p53, apoptosis and cell proliferation. Int J Cancer 112(3):393–398
26. Haybaeck J, O'Connor T, Spilka R, Spizzo G, Ensinger C, Mikuz G, Brunhuber T, Vogetseder A, Theurl I, Salvenmoser W, Draxl H, Banziger R, Bachmann F, Schafer G, Burger M, Obrist P (2010) Overexpression of p150, a part of the large subunit of the eukaryotic translation initiation factor 3, in colon cancer. Anticancer Res 30(4):1047–1055
27. Liu RY, Dong Z, Liu J, Yin JY, Zhou L, Wu X, Yang Y, Mo W, Huang W, Khoo SK, Chen J, Petillo D, Teh BT, Qian CN, Zhang JT (2011) Role of eIF3a in regulating cisplatin sensitivity and in translational control of nucleotide excision repair of nasopharyngeal carcinoma. Oncogene 30(48):4814–4823. doi:10.1038/onc.2011.189
28. Dong Z, Zhang JT (2003) EIF3 p170, a mediator of mimosine effect on protein synthesis and cell cycle progression. Mol Biol Cell 14(9):3942–3951
29. Lee AS, Kranzusch PJ, Cate JH (2015) eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature 522(7554):111–114. doi:10.1038/nature14267
30. Pradet-Balade B, Boulme F, Beug H, Mullner EW, Garcia-Sanz JA (2001) Translation control: bridging the gap between genomics and proteomics? Trends Biochem Sci 26(4):225–229, S0968-0004(00)01776-X [pii]
31. Yin JY, Dong ZZ, Liu RY, Chen J, Liu ZQ, Zhang JT (2013) Translational regulation of RPA2 via internal ribosomal entry site and by eIF3a. Carcinogenesis 34(6):1224–1231. doi:10.1093/carcin/bgt052

32. Han B, Dong Z, Liu Y, Chen Q, Hashimoto K, Zhang JT (2003) Regulation of constitutive expression of mouse PTEN by the 5′-untranslated region. Oncogene 22(34): 5325–5337
33. Liu Z, Dong Z, Han B, Yang Y, Liu Y, Zhang JT (2005) Regulation of expression by promoters versus internal ribosome entry site in the 5′-untranslated sequence of the human cyclin-dependent kinase inhibitor p27kip1. Nucleic Acids Res 33(12):3763–3771. doi:10.1093/nar/gki680, 33/12/3763 [pii]
34. Dong Z, Liu Y, Zhang JT (2005) Regulation of ribonucleotide reductase M2 expression by the upstream AUGs. Nucleic Acids Res 33(8):2715–2725
35. Han B, Zhang JT (2002) Regulation of gene expression by internal ribosome entry sites or cryptic promoters: the eIF4G story. Mol Cell Biol 22(21):7372–7384
36. Huang W, Dong Z, Chen Y, Wang F, Wang CJ, Peng H, He Y, Hangoc G, Pollok K, Sandusky G, Fu XY, Broxmeyer HE, Zhang ZY, Liu JY, Zhang JT (2015) Small-molecule inhibitors targeting the DNA-binding domain of STAT3 suppress tumor growth, metastasis and STAT3 target gene expression in vivo. Oncogene 35:783–792. doi:10.1038/onc.2015.215
37. Zhang Y, Yu JJ, Tian Y, Li ZZ, Zhang CY, Zhang SF, Cao LQ, Qian CY, Zhang W, Zhou HH, Yin JY, Liu ZQ (2015) eIF3a improve cisplatin sensitivity in ovarian cancer by regulating XPC and p27Kip1 translation. Onco target 6(28):25441–25451
38. Johannes G, Carter MS, Eisen MB, Brown PO, Sarnow P (1999) Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. Proc Natl Acad Sci U S A 96(23):13118–13123

Chapter 10

High-Resolution Gene Expression Profiling of RNA Synthesis, Processing, and Decay by Metabolic Labeling of Newly Transcribed RNA Using 4-Thiouridine

Andrzej J. Rutkowski and Lars Dölken

Abstract

Cellular RNA levels are orchestrated by highly regulated processes involving RNA synthesis (transcription), processing (e.g., splicing, polyadenylation, transport), and degradation. Profiling these changes provides valuable information on the regulation of gene expression. Total cellular RNA is a poor template for revealing short-term changes in gene expression, alterations in RNA decay rates, and the kinetics of RNA processing as well as the differentiation thereof. Here, we describe the metabolic labeling and purification of newly transcribed RNA with 4-thiouridine, by which these limitations are overcome.

Key words 4-Thiouridine, Biotin, Streptavidin, Newly transcribed RNA, RNA processing, RNA decay, Gene expression profiling, Microarray, RNA-seq

1 Introduction

Cellular RNA levels are subject to extensive regulation involving alterations in the rates of RNA synthesis (transcription), processing (e.g., splicing, polyadenylation, transport), and decay. Activation of cellular signaling pathways following both internal and external stimuli commonly results in substantial alterations in transcription rates [1]. However, changes in RNA degradation rates as well as impaired RNA processing may also significantly alter gene expression of both coding and noncoding RNAs [2–4]. High-throughput transcriptional analysis like microarrays or next-generation sequencing (RNA-seq) can provide important insights into the underlying molecular mechanism [5]. A major constraint of these analyses is the poor temporal resolution for kinetic changes. It is important to note that this is not a specific problem of these technologies but rather due to intrinsic properties of the involved biological samples. As such, total RNA levels of a transcript with an RNA half-life of 10 h will simply take 10 h to go down by twofold following a

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_10,

complete shutdown (e.g., >1000-fold downregulation) of transcription. The inability to differentiate changes in RNA synthesis rates from changes in RNA decay rates represents another major limitation. RNA decay rates can be studied by monitoring ongoing RNA decay over time following transcriptional arrest, e.g., by administrating actinomycin D [6]. However, this approach is not readily applicable to all models due to its cell-invasive nature and detrimental impact on host cell physiology. In addition, it relies on the detection of very small differences in total RNA levels and is thus inherently imprecise for more than half of all cellular genes [7]. Finally, the contribution of changes in RNA processing remains elusive. As such, while the alternatively spliced transcripts themselves are easily revealed by RNA-seq, the kinetics of this regulation and thus the underlying molecular mechanisms are not depicted.

All these problems can be overcome by metabolic labeling of newly transcribed RNA using 4-thiouridine (4sU-tagging; *see* Fig. 1). This approach provides direct access to newly synthesized

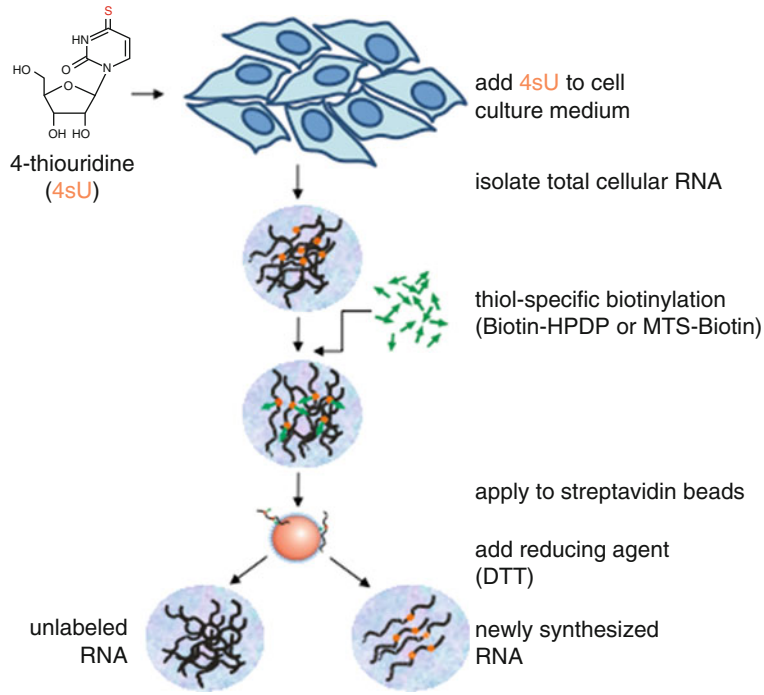

Fig. 1 Schematic overview on 4sU-tagging. Metabolic labeling of newly transcribed is initiated when 4sU is added to the cell culture medium. Following isolation of total cellular RNA, 4sU residues in newly transcribed RNA are thiol-specifically biotinylated. This allows separation of total RNA into newly transcribed (4sU-RNA) and unlabeled pre-existing RNA using streptavidin-coated magnetic beads. 4sU-RNA is recovered from the beads by adding a reducing agent, which cleaves the disulfide bond between 4sU and biotin. 4sU-RNA is recovered by column purification or isopropanol/ethanol precipitation

transcripts with minimal toxic effects [7–10]. Metabolic labeling is started by adding 4sU to cell culture medium and stopped by cell lysis followed by isolation of total cellular RNA. As RNA from mammalian cells does not contain thiol groups, the thiol-labeled newly transcribed RNA can be specifically biotinylated generating a disulfide bond between biotin residues and newly transcribed RNA molecules. This tag then allows rigorous purification of newly transcribed RNA using streptavidin-coated magnetic beads. Newly transcribed RNA is finally recovered from the beads by simply adding a reducing agent (dithiothreitol), which cleaves the disulfide bond and releases the newly transcribed RNA molecules from the beads. The purified newly transcribed RNA can then be directly subjected to qRT-PCR, microarray analysis, and RNA-seq. Employing this approach, snapshot pictures of the real-time kinetics of eukaryotic gene expression are obtained [11–13].

We recently applied this approach to study host cell modulation in lytic herpesvirus infections. During the first 6 h of lytic murine cytomegalovirus infection, discrete clusters of genes regulated with distinct kinetics were identified revealing the underlying molecular signaling events as well as viral counter-regulation thereof [14]. Applying both 4sU-tagging and ribosome profiling [15] to the full course of lytic herpes simplex virus 1 (HSV-1) infection, we elucidated detailed changes in RNA synthesis and processing as well as their impact on translation throughout infection [16]. Analysis of RNA molecules transcribed and processed at specific times of infection (defined by the exposure of the cells to 4sU) by RNA-seq revealed HSV-1 to disrupt transcription termination of cellular but not viral genes with dramatic effects on cellular gene expression. In addition, we surprisingly found that splicing, which according to textbook knowledge is globally inhibited by HSV-1, remains fully intact throughout infection. In contrast, poly(A) read-through and transcription into downstream genes were associated with a failure in splicing. These findings highlight the benefits of profiling newly transcribed rather than total RNA.

Under steady-state conditions, RNA synthesis compensates for ongoing RNA decay. Therefore, the ratios of newly transcribed/total or newly transcribed/unlabeled RNA allow measurements of RNA half-lives [12, 17, 18]. These are much more precise than RNA half-lives obtained by transcriptional arrest [17]. When all three RNA fractions, i.e. newly transcribed, total and unlabeled, pre-existing RNA, are analyzed, both data normalization and data quality control are intrinsically provided by a linear regression analysis model [17]. In case of non-steady-state conditions, comprehensive analysis of the changes in transcription rates and total RNA levels allows computational modeling of kinetic changes in RNA synthesis and decay [11, 13]. Finally, it is important to note that the mean age of newly transcribed RNA (4sU-RNA) depends on the duration of 4sU exposure (i.e., it increases with the duration of

4sU-labeling). Therefore, when 4sU-RNA is isolated following different durations of labeling and then analyzed by RNA-seq (*see* **Note 1**), the kinetics of RNA processing are revealed at nucleotide resolution [19]. In conclusion, 4sU-tagging provides access to the dynamics of RNA synthesis, processing, and decay in eukaryotic cells including all major model organisms [7, 11, 20].

In this chapter, the methodology to metabolically label, biotinylate, and purify newly transcribed RNA from total cellular RNA is described. In all our previous work, we employed Biotin-HPDP to biotinylate the 4sU residues in newly transcribed RNA. Biotin-HPDP is a 100% thiol-selective, reversible biotinylation agent. It results in the formation of a disulfide bond between 4sU and biotin, allowing subsequent cleavage and release of unbiotinylated RNA from the streptavidin beads using a reducing agent (e.g., dithiothreitol). In addition, Biotin-HPDP is not water soluble and can thus be efficiently removed from the RNA samples by chloroform extraction. In contrast to more highly reactive biotinylation reagents like iodoacetyl-biotin, Biotin-HPDP only biotinylates about one in three 4sU residues in 4sU-labeled RNA [7]. A recent report shows that more efficient biotinylation can be achieved with methanethiosulfonate (MTS)-biotin, which, just like Biotin-HPDP, is reversible and water insoluble. However, it is not 100 % thiol-specific (unpublished observations) and may thus cause problems with unspecific biotinylation of RNA, particularly when rather small amounts of RNA are to be purified (ultra-short 4sU-labeling). Nevertheless, 4sU-tagging performed using MTS-biotin now allows studies on RNA synthesis and turnover of smaller RNA species (e.g., miRNAs) [21]. In order to compensate for the greater amount of biotin residues introduced, when using MTS-biotin, the volume of streptavidin beads mentioned in this protocol should be doubled [21]. As further optimisation is still required for using MTS-biotin, this protocol only describes the use of Biotin-HPDP.

2 Materials

Prepare all solutions using nuclease-free water. Using in-house purified, deionized water can result in problems due to the presence of trace amounts of RNases or reducing agents and might lead to the complete loss of labeled RNA. Therefore, we strongly recommend buying commercially available nuclease-free NaCl, Tris–HCl, EDTA, and water. Take all necessary precautions to ensure nuclease-free conditions at all times. All solutions/reagents should be stored at room temperature unless otherwise specified.

2.1 Metabolic Labeling and Isolation of Total Cellular RNA

1. 4-thiouridine: dissolve in sterile water to 50 mM stock concentration, store in small aliquots of 50–500 µl at −20 °C, discard unused reagent after use (do not refreeze).

2. TRIzol, store at 4 °C.
3. 15 ml polypropylene tubes SuperClear™ GateFree™, in contrast to standard Falcon tubes, these tolerate up to 20,000 × *g*.
4. Polypropylen adaptors for four 15 ml "Falcon" tubes 62 × 120 mm.
5. Chloroform.
6. Isopropanol.
7. DEPC (to inactivate RNases), store at 4 °C.
8. Sodium citrate: make 1.6 M stock solution, treat with 0.1% DEPC overnight with shaking to eliminate RNases and autoclave thereafter, alternatively buy nuclease-free sodium citrate, and dissolve using nuclease-free water.
9. 5 M nuclease-free NaCl.
10. RNA precipitation buffer: 1.2 M NaCl, 0.8 M sodium citrate.
11. Ethanol.
12. Nuclease-free H_2O.

2.2 Biotinylation of Thiol-Labeled, Newly Transcribed RNA

1. 1 M nuclease-free Tris–HCl, pH 7.5.
2. 500 mM nuclease-free EDTA, pH 8.0.
3. Nuclease-free 10× biotinylation buffer (BB): 100 mM Tris–HCl, pH 7.4, 10 mM EDTA, store in aliquots of 1–1.5 ml at 4 °C.
4. Dimethylformamide.
5. EZ-Link Biotin-HPDP: 1 mg/ml stock concentration dissolved in dimethylformamide, gentle warming will ensure complete solubilization, store in aliquots of 1 ml at 4 °C (*see* **Note 2**).
6. Nuclease-free H_2O.
7. Chloroform.
8. Phase Lock Gel (2.0 ml) Heavy Tubes.
9. Isopropanol.
10. Ethanol.

2.3 Streptavidin Capture

1. μMacs Streptavidin Kit (Miltenyi, *see* **Note 3**).
2. Magnetic Stand (Miltenyi, one stand holds four or eight columns of the μMacs Streptavidin Kit).
3. Tween 20.
4. Washing buffer (WB): 100 mM, Tris–HCl, pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween 20.
5. Dithiothreitol (DTT): 100 mM DTT in nuclease-free H_2O, always prepare fresh before use.
6. RNeasy MinElute Kit, store columns at 4 °C.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Metabolic Labeling of Newly Transcribed RNA with 4-Thiouridine

As 4-thiouridine is very efficiently taken up by cells, using too little volume of cell culture media may reduce labeling efficiency. We therefore recommend using 5 ml/10 ml of medium for a 10 cm/15 cm dish, respectively. Before beginning the labeling, make a detailed time plan of the whole experiment allowing for 5 min between each condition (usually comprising up to five technical replicates).

3.1.1 Beginning of Labeling

1. Thaw 4-thiouridine (4sU) just before use, and pipette required amount of 4sU for each condition into a sterile Falcon tube.
2. Only treat cells of one condition at a time. Try to handle cells as quickly as possible.
3. Take the required amount of medium off the plates, add this to the 4sU-containing Falcon tube, and mix well. Discard the remaining medium from the plates.
4. Reapply 4sU-containing medium back to cell culture plates.

3.1.2 End of Labeling

1. Carefully remove all cell culture medium from cells (one condition at a time, max. 3–5 plates), and immediately add TRIzol to each plate (5 ml per 15 cm dish). For complex experiments including multiple time points, this step is best done by two people, one removing the medium, the other adding TRIzol and harvesting the lysate.
2. Pipette up and down several times and incubate at room temperature for 5 min to facilitate complete cell lysis.
3. Transfer to polypropylene tubes. Samples can be stored at −20 °C for up to 1 month until RNA is prepared.

3.2 RNA Preparation Using Modified TRIzol Protocol

To ensure complete removal of unlabeled RNA, it is important to obtain very clean total cellular RNA (*see* **Note 4**). Therefore, do not use too little lysis reagent. In case of in vivo samples, preparation of poly-A RNA may be required.

1. Add 1 ml chloroform (0.2 ml per 1 ml TRIzol) and shake vigorously for 15 s. Incubate at room temperature for 2–3 min.
2. Centrifuge at 13,000 × *g* for 15 min at 4 °C.
3. Transfer aqueous upper phase (containing the RNA) to a new 15 ml polypropylene tube.
4. Add ½ volume of both RNA precipitation buffer and isopropanol (e.g., to 3 ml supernatant aqueous phase, add 1.5 ml RNA precipitation buffer and 1.5 ml isopropanol).

5. Mix well. Incubate at room temperature for 10 min.
6. Centrifuge at 13,000 × *g* for 10 min at 4 °C. Discard supernatant.
7. Spin down briefly and remove residual isopropanol with 200 µl pipette.
8. Add an equal volume of 75 % ethanol and shake carefully until the pellet detaches. Avoid extensive vortexing as disrupting the pellet too much may make the removal of residual ethanol challenging.
9. Centrifuge at 13,000 × *g* for 10 min at 4 °C. Immediately discard supernatant. Spin down briefly and remove remaining ethanol with a 200 µl pipette. No further drying of the pellet is necessary.
10. Add 100 µl of H_2O per 100 µg expected RNA yield and transfer to RNase-free microcentrifuge tubes.
11. Dissolve RNA by heating to 65 °C for 10 min with shaking and immediately place on ice.
12. Measure RNA spectrophotometrically. This RNA can be stored at −80 °C for at least 1 month.

3.3 Biotinylation

Carry out all procedures at room temperature. Avoid direct exposure to bright light as 4sU is light sensitive. Use 30–150 µg total RNA for biotinylation.

Labeling reaction (per 100 µg RNA; pipette in this order):

1. 100 µl 10× biotinylation buffer
2. 100 µg RNA
3. Water up to 800 µl
4. 200 µl Biotin-HPDP (1 mg/ml DMF)
   Always add the Biotin-HPDP last and mix immediately by pipetting. In case the biotin precipitates, DMF content can be increased to a final concentration of 40 %.

1. Incubate at room temperature for 1.5 h with rotation.
2. Add an equal volume of chloroform. Mix vigorously. Incubate for 2–3 min until phases begin to separate and bubbles start to disappear (*see* **Note 5**).
3. Centrifuge at full speed (20,000 × *g*) for 5 min. Carefully transfer upper phase into new tubes.
4. Repeat **steps 2** and **3** once. This step can also be performed using Phase Lock Gel (2.0 ml) Heavy Tubes.
5. RNA precipitation: add 1/10 the volume of 5 M NaCl and an equal volume of isopropanol.
6. Centrifuge at 20,000 × *g* for 20 min at 4 °C. Discard supernatant.

7. Add an equal volume of 75% ethanol, centrifuge at 20,000 × *g* for 10 min, and discard supernatant.
8. Spin briefly and remove residual ethanol by pipetting as described above.
9. Do not allow RNA to dry but immediately resuspend it in 100 µl H_2O.

3.4 Separation of Labeled and Unlabeled RNA Using Streptavidin-Coated Magnetic Beads

1. Heat up washing buffer to 65 °C in a 50 ml Falcon tube (3 ml per sample).
2. Prepare fresh 100 mM dithiothreitol (DTT) in nuclease-free H_2O for elution.
3. Heat biotinylated RNA samples to 65 °C for 10 min to denature and immediately place on ice for 5 min.
4. Pre-equilibrate Miltenyi columns with 1 ml room temperature washing buffer. This will take about 15 min.
5. Add 100 µl of biotinylated RNA to 100 µl of streptavidin beads. Incubate with rotation for 15 min.
6. Place µMacs columns into magnetic stand. Do not process more than 12 samples at a time (six to eight samples are optimal).
7. Apply beads (RNA) to the columns. Discard the flow-through (unless unlabeled RNA is of interest; *see* Subheading 3.6).
8. Wash 3× with 0.9 ml 65 °C washing buffer (actual aspirated volumes can be larger than the volume set on an automatic pipette when handling hot solutions).
9. Wash 3× with 0.9 ml room temperature washing buffer.
10. Pipette 700 µl Buffer RLT (RNeasy MinElute Cleanup Kit) into new 2 ml tubes.
11. Elute RNA directly into Buffer RLT by placing the tubes underneath the columns and adding 100 µl 100 mM DTT to the columns.
12. Perform a second elution round into the same tubes 3 min later.

3.5 Recovery of Newly Transcribed RNA

Continue with the RNeasy MinElute (Qiagen) cleanup protocol following the manufacturer's instructions shown below (*see* **Note 6**).

1. Add 500 µl 96–100% ethanol to the diluted RNA and mix thoroughly by pipetting. Do not centrifuge.
2. Apply 700 µl of the sample to an RNeasy MinElute spin column in a 2 ml collection tube. Close the tube gently and centrifuge for 15 s at >8000 × *g*. Discard the flow-through.
3. Apply the remaining 700 µl and repeat the centrifugation. Discard the flow-through. Transfer the spin column into a new 2 ml collection tube.

4. Pipette 500 µl Buffer RPE onto the spin column. Close the tube gently and centrifuge for 15 s at >8000 × *g* to wash the column. Discard the flow-through. Transfer the spin column into a new 2 ml collection tube (not supplied).
5. Add 500 µl of 80% ethanol to the spin column. Close the tube gently and centrifuge for 2 min at >8000 × *g* to dry the silica gel membrane. Discard the flow-through and transfer the spin column into a new 2 ml collection tube. Open the cap of the spin column and centrifuge at full speed for 5 min.
6. Transfer the spin column to a new 1.5 ml collection tube. Pipette 20 µl nuclease-free water directly onto the center of the silica gel membrane. Close the tube gently and incubate for 1 min before centrifuging for 1 min at maximum speed to elute.
7. Measure RNA concentration using a NanoDrop 1000 Spectrophotometer.
8. For qRT-PCR analysis we recommend to use 2.5 µl of labeled RNA in 20 µl cDNA synthesis mix. Subject 1:10 dilutions to qRT-PCR.
9. Store RNA at −80 °C.

3.6 Recovery of Unlabeled, Unbound RNA

In case the unlabeled (pre-existing) RNA needs to be recovered, collect the flow-through and the first wash (Subheading 3.4, **steps 7** and **8**) for subsequent precipitation (together these contain >90% of the unbound RNA). Combine the two fractions and recover the unbound RNA by isopropanol/ethanol precipitation as performed after the biotinylation reaction (no salt needs to be added as the washing buffer already contains 1 M NaCl).

4 Notes

1. For RNA-seq, rRNA depletion is of much lesser importance as newly transcribed RNA contains substantially less rRNA than total RNA (50–60% compared to ≈95%). Therefore, the gain in sequencing depth achieved by complete rRNA depletion is not nearly as great as for total RNA (2-fold compared to 20-fold).
2. It is crucial to prevent the dimethylformamide from getting in contact with incompatible plastic materials. Otherwise, substances are eluted from the plasticware and carried along through the chloroform extraction and isopropanol/ethanol precipitation steps, which cause a substantial loss (>75%) of newly transcribed RNA during the streptavidin capture. We believe this to be due to damage to the coating of the Miltenyi beads. The problem becomes even more prominent when the duration of labeling is shortened to 30 min or less. The same problem may occur when cell scrapers are used to collect the TRIzol samples from cell culture plates. This should thus be avoided.

3. We would strongly recommend not changing the provider of the streptavidin beads as we tested beads from four different companies and found only these to work without background RNA carry-over. The column-based purification allows for easy handling.
4. Using this modified TRIzol protocol by Chomczynski et al. [22] improves the removal of DNA and glycoproteins. 5 ml TRIzol per 15 cm dish produces clean RNA. Reducing the amount of TRIzol may result in incomplete removal of RNases and subsequent RNA degradation. As higher centrifugal forces are used, the RNA pellets are more solid and easier to handle. This requires the use of special polypropylene tubes as the regular 15 ml Falcon tubes do not survive more than 6000 × *g*.
5. The chloroform extraction is required to remove unincorporated Biotin-HPDP. To reduce template RNA loss during the chloroform extraction, step Phase Lock Gel Heavy tubes (2.0 ml, Eppendorf) may be used following the manufacturer's instructions. We only use the phase-lock tubes for the second chloroform extraction as >1 ml biotinylation volume is too much volume to fit into these tubes.
6. Recovery of newly transcribed RNA is highly quantitative. If you started with the same RNA concentration, you can expect the same amounts of newly transcribed RNA. In case the yields of labeled RNA are lower than expected, carefully look for signs of RNA degradation by electrophoretic analysis of the newly transcribed RNA. Newly transcribed RNA is of higher molecular weight than total cellular RNA due to the presence of large unspliced transcripts and a slightly reduced contribution of rRNA. $OD_{260/280}$ ratios below 1.7 (instead of ~2.0) usually indicate the carry-over of washing buffer from the Qiagen kit. While this does not pose a major problem in downstream analysis (e.g., cDNA synthesis), it confounds RNA measurements (overestimating the amount of purified RNA). This problem is reduced by changing the collection tube after each centrifugation step in the RNeasy MinElute collection step or by precipitating the newly transcribed RNA with glycogen and isopropanol/ethanol.

Acknowledgment

This work was supported by MRC fellowship grant G1002523 and NHSBT grant WP11-05 to L.D.

References

1. Lee TI, Young RA (2013) Transcriptional regulation and its misregulation in disease. Cell 152(6):1237–1251. doi:10.1016/j.cell.2013.02.014
2. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11(1):75–87. doi:10.1038/nrg2673
3. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36(10):1073–1078. doi:10.1038/ng1429, ng1429 [pii]
4. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505(7483):344–352. doi:10.1038/nature12986
5. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63
6. Yang E, van Nimwegen E, Zavolan M, Rajewsky N, Schroeder M, Magnasco M, Darnell JE Jr (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res 13(8):1863–1872
7. Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, Koszinowski UH (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA 14(9):1959–1972
8. Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM (1978) Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur J Biochem 92(2):373–379
9. Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol 23(2):232–237
10. Kenzelmann M, Maertens S, Hergenhahn M, Kueffer S, Hotz-Wagenblatt A, Li L, Wang S, Ittrich C, Lemberger T, Arribas R, Jonnakuty S, Hollstein MC, Schmid W, Gretz N, Grone HJ, Schutz G (2007) Microarray analysis of newly synthesized RNA in cells and animals. Proc Natl Acad Sci U S A 104(15):6164–6169
11. Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L, Martin DE, Tresch A, Cramer P (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol 7:458
12. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature 473(7347):337–342. doi:10.1038/nature10098, nature10098 [pii]
13. Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, Amit I, Regev A (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat Biotechnol 29(5):436–442. doi:10.1038/nbt.1861
14. Marcinowski L, Lidschreiber M, Windhager L, Rieder M, Bosse JB, Radle B, Bonfert T, Gyory I, de Graaf M, Prazeres da Costa O, Rosenstiel P, Friedel CC, Zimmer R, Ruzsics Z, Dolken L (2012) Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection. PLoS Pathog 8(9), e1002908. doi:10.1371/journal.ppat.1002908, PPATHOGENS-D-12-00468 [pii]
15. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324(5924):218–223
16. Rutkowski AJ, Erhard F, L'Hernault A, Bonfert T, Schilhabel M, Crump C, Rosenstiel P, Efstathiou S, Zimmer R, Friedel CC, Dolken L (2015) Widespread disruption of host transcription termination in HSV-1 infection. Nat Commun 6:7126. doi:10.1038/ncomms8126
17. Friedel CC, Dolken L, Ruzsics Z, Koszinowski UH, Zimmer R (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res 37(17), e115. doi:10.1093/nar/gkp542, gkp542 [pii]
18. Dolken L, Malterer G, Erhard F, Kothe S, Friedel CC, Suffert G, Marcinowski L, Motsch N, Barth S, Beitzinger M, Lieber D, Bailer SM, Hoffmann R, Ruzsics Z, Kremmer E, Pfeffer S, Zimmer R, Koszinowski UH, Grasser F, Meister G, Haas J (2010) Systematic analysis of viral and cellular microRNA targets in cells latently infected with human gammaherpesviruses by RISC immunoprecipitation assay. Cell Host Microbe 7(4):324–334
19. Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, Malterer G, L'Hernault A, Schilhabel M, Schreiber S, Rosenstiel P, Zimmer R, Eick D, Friedel CC, Dolken L (2012) Ultrashort and progressive 4sU-tagging reveals

key characteristics of RNA processing at nucleotide resolution. Genome Res 22(10):2031–2042. doi:10.1101/gr.131847.111

20. Miller MR, Robinson KJ, Cleary MD, Doe CQ (2009) TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat Methods 6(6):439–441
21. Duffy EE, Rutenberg-Schoenberg M, Stark CD, Kitchen RR, Gerstein MB, Simon MD (2015) Tracking distinct RNA populations using efficient and reversible covalent chemistry. Mol Cell 59(5):858–866. doi:10.1016/j.molcel.2015.07.023
22. Chomczynski P, Mackey K (1995) Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques 19(6):942–945

Chapter 11

Accurate Detection of Differential Expression and Splicing Using Low-Level Features

Tomi Suomi and Laura L. Elo

Abstract

Gene expression can be quantified in high throughput using microarray technology. Here we describe how to accurately detect differential expression and splicing using a probe-level expression change averaging (PECA) method. PECA is available as an R package from Bioconductor (https://www.bioconductor.org), and it supports multiple operating systems.

Key words PECA, Differential expression, Differential splicing, Probe level

1 Introduction

Microarray technology is a high-throughput method to quantify gene expression on thousands of genes simultaneously. Microarrays contain probe sequences that represent various genes to which the cRNAs in the sample bind. The tested samples are labeled with fluorescent dyes beforehand, and the intensities of the fluorescent emissions in different parts of the microarray can be read using a microarray scanner, thus allowing to determine the gene expression in the sample. PECA is specifically developed for the analysis of measurements from the popular microarray platform Affymetrix, which utilizes multiple different probes to measure single genes (called probe sets).

Multiple methods and improvements for detecting differentially expresses genes have been proposed. Cui and Churchill review some widely used methods and their modifications including the *t*-test and analysis of variance (ANOVA) [1]. Huber et al. propose a method to apply variance-stabilizing transformation to gene expression experiments to counter the dependence of the standard deviation of an intensity on its mean [2]. Many of the proposed methods are also available as software such as the popular limma package [3].

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_11,

The improved performance of using low-level measurements when detecting differential gene expression has been demonstrated earlier by us and others [4–7]. For instance, Liu et al. extended the modified gamma Model for Oligonucleotide Signal (mgMOS) to model the binding affinity of probe pairs across multiple microarrays [8] and proposed a method that includes probe-level measurement errors from replicated experiments in calculating differential expression of a gene [9]. There are also methods such as Logit-t [10] and ChipStat [11], which perform *t*-tests on probe level. An online algorithm for efficient use of probe-level data on preprocessing has also been proposed [12]. The method shown here, named PECA, determines differential gene expression using directly the probe-level values from Affymetrix gene expression microarrays. The probe-level expression changes are calculated using the ordinary or modified *t*-statistic, and the median *t*-statistics are used to calculate gene-level scores. Finally, the gene-level significances of *p*-values are calculated from the beta distribution [13] or by using simulated distributions.

Similarly as for detecting differential gene expression, we have shown the benefits of using probe-level data to detect differential splicing [14]. Alternative splicing is a process that allows a single gene to code multiple proteins. This happens when mRNA is formed from a gene, and its particular exons are included or excluded in different patterns during the process. Proteins that are translated from these mRNA isoforms have differences in their amino acid sequence. For this purpose there are exon microarrays available that have multiple probes per exon and allow to distinguish the different isoforms when quantifying gene expressions. It has been estimated that approximately 95% of multiexon human genes undergo alternative splicing [15]. For detecting differential splicing between sample groups, one widely used approach is splicing index (SI), where the exon-level intensities are first normalized by the corresponding gene-level values and then those normalized values are compared between the sample groups [16]. Other methods include ANOSVA [17] and MIDAS [18], both of which are based on analysis of variance.

Here we show practical examples using probe-level expression change averaging (PECA) method when calculating differential gene expression and differential splicing.

2 Materials

The following examples are run using R environment for statistical computing which supports multiple operating systems. It can be freely obtained from http://www.r-project.org, where installation

instructions and basic guides are also available. The PECA package is available from Bioconductor (https://www.bioconductor.org), which is an open source project providing tools for genomic data analysis and is mainly based on R programming language. To install PECA, open R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("PECA")
```

The following command also installs spike-in microarray dataset that is used in the example of Section 3.1. The spike-in data includes AffyBatch objects containing the perfect match (PM) and mismatch (MM) intensities for genes spiked in at different concentrations on Affymetrix HGU95 and HGU133 spike-in experiments. The spike-in experiments allow us to evaluate the performance of different methods in terms of their accuracy because the changes between different sample groups are known beforehand. They contain a number of spiked genes and some background that is unchanged between the groups. By using this information, we are able to determine the number of true positives and false positives when looking at the results. The improved accuracy of PECA on HGU133 experiment is shown in **Note 1**.

```
source("http://bioconductor.org/biocLite.R")
biocLite("SpikeIn")
```

Human Exon 1.0 ST array data that is used in differential splicing analysis in the example of Section 3.3 can be downloaded from Affymetrix (http://www.affymetrix.com/support/technical/sample_data/gene_1_0_array_data.affx).

3 Methods

The following examples are run on R version 3.2.0 *Full of Ingredients* and PECA version 1.4.0. Development version of PECA 1.5.2 was used in Affymetrix exon array examples.

3.1 Differential Gene Expression Analysis of Built in Spike-In Dataset

In this section we show step-by-step instructions on how to perform differential gene expression analysis using spike-in data available from Bioconductor. This allows inexperienced users to get familiar with the R environment, and it requires nothing more than copying the listed command into the R console:

1. Open R and load the PECA package along with the spike-in example dataset.
   ```
   library(PECA)
   library(SpikeIn)
   ```
2. Load the example dataset into memory and subset only part of the data for analysis for convenience. The parts of this Latin

square data, which are selected for differential expression analysis, are Expt1_R1 to R3 and Expt2_R1 to R3. Both of the experiments contain three replicates, and in most cases the spike-in concentrations have twofold increase. For full details, see the *SpikeIn133* documentation in R/Bioconductor (https://bioconductor.org).

```
data(SpikeIn133)
data <- SpikeIn133[,c(1,15,29,2,16,30)]
```

3. Run the differential expression analysis using PECA. The AffyBatch object data, that was subsetted above is given as a parameter to the PECA function. Data normalization is requested so quantile normalization is performed as a default option. Other options are also available (*see* **Note 2**). The function prints out information of current processing steps while the tests are being performed, and in this case the results are being stored to `peca_results`.

```
peca_results <- PECA_AffyBatch(affy=data,
normalize=TRUE)
```

4. The resulting data frame `peca_results` contains all the information that was calculated during the analysis, and it can be processed further. The following command can be used to see all genes that have their false discovery rate (FDR) < 0.05 (*see* **Notes 3** and **4**). Additional visualization is possible by plotting the signal log ratios of genes (*see* **Note 5**) or by using volcano plots (*see* **Note 6**).

```
peca_results[peca_results$p.fdr < 0.05,]
```

5. The write.table command in R can be used to store data frames and matrices to disk. The following command can be used to write the results into current R working directory using results.txt as the file name and tabulator as a separator for columns. Storage place can be changed either by setting a different working directory or by including the full path in the file name. Results can also be sorted beforehand (*see* **Note 7**).

```
write.table(peca_results, file="results.txt",
sep="\t")
```

3.2 Differential Gene Expression Analysis of Affymetrix .CEL Files

In this section we show step-by-step instructions on how to perform differential gene expression analysis as earlier but by using .CEL files provided by the users themselves.

1. Open R and load the PECA package.

```
library(PECA)
```

2. Set the file names of .CEL files divided into desired groups for differential expression analysis.

```
group1 <- c("Exp_A_1.CEL", "Exp_A_2.CEL",
"Exp_A_3.CEL")
```

```
group1 <- c("Exp_B_1.CEL", "Exp_B_2.CEL",
"Exp_B_3.CEL")
```

3. Run the differential expression analysis using PECA. The file names are taken from vectors group1 and group2, which are given as parameters. Data normalization is requested so quantile normalization is performed as a default option. Other options are also available (*see* **Note 2**). This example requires that the .CEL files are in R working directory. Either set the correct working directory using the `setwd` command or define full paths when setting the groups. PECA also prints out information of current processing steps while the tests are being performed, and in this case the results are being stored to `peca_results`.
```
peca_results <- PECA_CEL(group1, group2,
normalize=TRUE)
```
4. The resulting data frame contains all the information that was calculated during the analysis, and it can be processed further. The following command can be used to see all the probes that have their false discovery rate (FDR) < 0.05 (*see* **Notes 3** and **4**). Additional visualization is possible by plotting the signal log ratios of genes (*see* **Note 5**) or by using volcano plots (*see* **Note 6**).
```
peca_results[peca_results$p.fdr < 0.05,]
```
5. The write.table command in R can be used to store data frames and matrices to disk. To write the results into current R working directory using results.txt as the file name and tabulator as a separator for columns, the following command can be used. Storage place can be changed either by setting a different working directory or by including the full path in file name. Results can also be sorted beforehand (*see* **Note 7**).
```
write.table(peca_results, file="results.txt",
sep="\t")
```

3.3 Differential Splicing Analysis of Affymetrix Exon .CEL Files

In this section we show step-by-step instructions on how to perform differential splicing analysis by using CEL files. The example files are available from Affymetrix. Users can easily use these steps to analyze their own data as well.

PECA contains PECASI function that uses the aroma.affymetrix package to normalize and extract the probe-level data from the .CEL files [19]. Therefore, it is important that the naming and structure of the data files follow exactly the rules specified in the aroma.affymetrix package.

The raw expression data (.CEL files) need to be in the directory rawData/[dataFolder]/[chipType], where rawData is a directory under the current working directory specified by the path shown later, dataFolder is the name of the dataset given by the user, and chipType indicates the type of the microarray used in the experiment.

In addition to the expression data, a chip definition file (CDF) is required. The CDF file(s) for a particular microarray type chipType need to be in the directory annotationData/chipTypes/[chipType], where annotationData is a directory under the current working directory specified by the path. Besides the CDF files provided by Affymetrix, various custom CDF files are available for a particular microarray type. The different versions can be separated by adding a suffix cdfTag to the name of the CDF file: [chipType],[cdfTag].cdf.

1. Set the file names of .CEL files divided into desired groups for differential splicing analysis. Names should be given without the .CEL extension in file name.
```
group1 <- c("TisMix_mix2_01_v1_Exon1",
   "TisMix_mix2_02_v1_Exon1", "TisMix_
mix2_03_v1_Exon1")
group2 <- c("TisMix_mix8_01_v1_Exon1",
   "TisMix_mix8_02_v1_Exon1", "TisMix_
mix8_03_v1_Exon1")
```
2. Run the differential splicing analysis using PECASI. The file names are taken from vectors group1 and group2, which are given as parameters. PECASI prints out information of current processing steps while the tests are being performed, and in this case the results are being stored to pecasi_results. Users are encouraged to follow the filtering steps described by Affymetrix (*see* **Note 8**).
```
pecasi_results <- PECASI(path="[your direc-
tory path]",
   cdfTag="U-Ensembl49,G-Affy",
   dataFolder="TisMix",
   chipType="HuEx-1_0-st-v2",
   samplenames1=group1, samplenames2=group2,
   test="t")
```
3. The following command can be used to show the results that have their estimated *p*-values under 0.01.
```
pecasi_results[pecasi_results$p < 0.01,]
```
4. The write.table command in R can be used to store data frames and matrices to disk. To write these results into current R working directory using results.txt as the file name and tabulator as a separator for columns, the following command can be used. Storage place can be changed either by setting a different working directory or by including the full path in file name. Results can also be sorted beforehand (*see* **Note 7**).

   write.table(pecasi_results, file="results.txt", sep="\t")

4 Notes

1. The receiver operating characteristic (ROC) curves can be used to assess the sensitivity and specificity of different models (Fig. 1). They are created by plotting the true positive rate (sensitivity) against the false positive rate (1 - specificity) using various thresholds. The ROC curve in Figure 1 shows the performance of PECA on Affymetrix HGU133 SpikeIn experiment when using the probe-level values or summarized gene-level values as an input for differential expression analysis. Both the *x*-axis and the *y*-axis are scaled from 0.6 to 1.0 as this is the most interesting region for practical purposes. The ROC curve shows that when PECA is used with probe-level values, its true positive rate increases faster than that of gene-level method in the region of interest.
2. PECA has some default values that are used as settings in various processing steps. The following list contains all possible parameters, which can be changed, together with their default values.

   (a) Affy: AffyBatch object containing intensity data.

   (b) Normalize: Character string ("quantile", "median") or logical indicating the type of normalization. Default is false; no normalization is performed. True performs quantile normalization.

   (c) Test: Character string indicating whether the ordinary "t" or modified "modt" *t*-test is performed. Default is t.

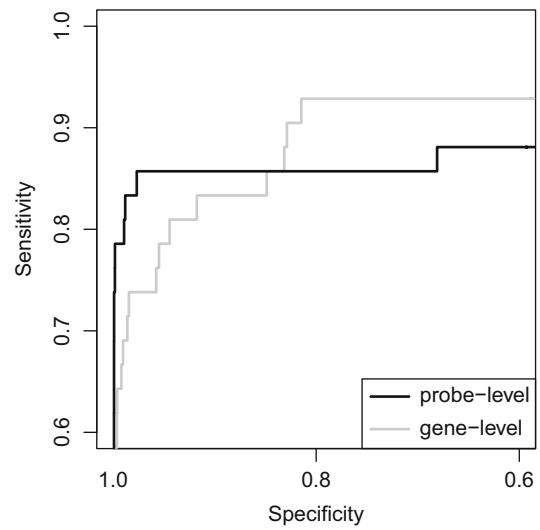

Fig. 1 ROC. Receiver operating characteristic (ROC) curve of PECA on Affymetrix HGU133 SpikeIn experiment when using the probe-level values or summarized gene-level values as an input for differential expression analysis

(d) Type: Character string indicating whether "median" or "Tukey" is used when calculating gene-level values. Default is median.

(e) Paired: Logical indicating whether a paired test is performed. Default is false; unpaired test is performed.

(f) Samplenames1: Character vector containing the names of the .CEL files in the first group.

(g) Samplenames2: Character vector containing the names of the .CEL files in the second group.

3. When looking at the statistical significance of median *p*-value for each gene, the corresponding *p*-values are determined from the beta distribution [13]. Under the null hypothesis, the *p*-values of the *n* probes that correspond to a gene will follow the uniform distribution $U(0,1)$ and furthermore the order statistics from that distribution have beta distributions. Additionally, the adjusted *p*-values for multiple testing are reported. These are calculated using the Benjamini-Hochberg method [20].
4. PECA provides an option to aggregate the results using Tukey's biweight instead of median and then the statistical significance of each gene is based on a simulated distribution. This is created by repeatedly storing Tukey's biweight values from a set of random *p*-values, which are based on the total number of probes in the given gene. Additionally, the adjusted *p*-values for multiple testing are reported. These are calculated using the Benjamini-Hochberg method [20].
5. Results can be visualized using a bar plot of signal log ratios from most differentially expressed items (Fig. 2). Here the threshold for filtering is *p*-value 0.05. Bars below the *x*-axis depict genes that are more expressed in sample group2, and

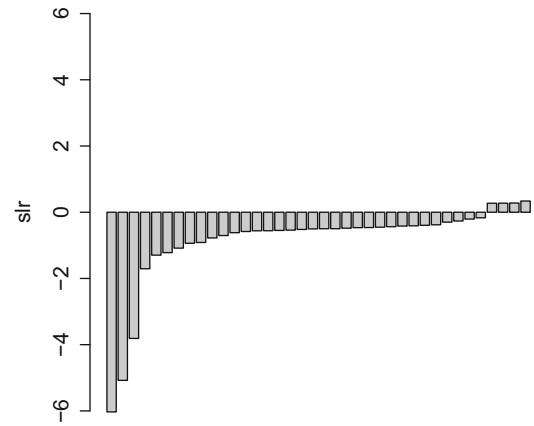

Fig. 2 Bar plot. Bar plot of signal log ratios (sir) from most differentially expressed items

bars above the *x*-axis are the ones expressed more in sample group1. One unit of change in signal log ratio means that the fold change in expressions between the groups has doubled. This does not provide any additional information compared to the result table, but it provides a fast and convenient way to see direction and magnitude of differential expression.

```
peca_results <- peca_results[peca_results$p
< 0.05,]
max_slr <- max(abs(peca_results$slr),
na.rm=TRUE)
barplot(peca_results$slr[order(peca_
results$slr)],
ylim=c(-max_slr, max_slr),
ylab="slr")
```

6. Results can be visualized using volcano plot, a comparison between *p*-values and signal log ratios for each item (Fig. 3). It shows the relationship between the magnitude of change and the significance of statistical test. This can be used, for instance, to select the most promising candidates for further validation.

```
slr <- peca_results$slr
p <- peca_results$p
max_slr <- max(abs(slr), na.rm=TRUE)
max_p <- max(-log10(p), na.rm=TRUE)
plot(slr, -log10(p),
   xlim=c(-max_slr, max_slr), ylim=c(0,
max_p),
   xlab="slr", ylab="-log10 p-value",
   pch=16)
```

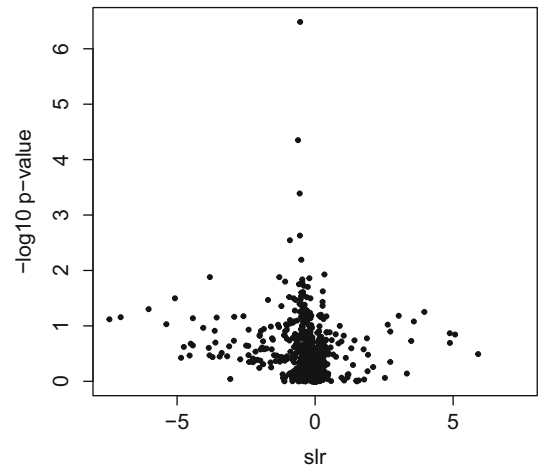

Fig. 3 Volcano plot. Volcano plot showing comparison between *p*-values and signal log ratios (slr) for each item

7. Sorting the results (e.g., based on *p*-values) can be performed before writing results to a file.
```
peca_results <- peca_results[order(peca_
results$p),]
```
8. When using Affymetrix exon arrays for alternative splicing analysis, users are encouraged to follow the filtering steps presented in detail in Affymetrix technical note [21]. The steps that are stated mandatory are removing the transcript clusters (genes) that are not expressed in both sample groups and also removing the probe sets (exons) that are not expressed at least in one sample group. It is also suggested to require minimum signal level for genes and to remove probe sets with either very large or very low exon/gene intensity ratios.

References

1. Cui X, Churchill GA (2003) Statistical tests for differential expression in cDNA microarray experiments. Genome Biol 4(4):210
2. Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18(suppl 1):S96–S104
3. Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) Bioinformatics and computational biology solutions using R and Bioconductor. Springer, New York, pp 397–420
4. Elo LL, Lahti L, Skottman H, Kyläniemi M, Lahesmaa R, Aittokallio T (2005) Integrating probe-level expression changes across generations of Affymetrix arrays. Nucleic Acids Res 33(22):e193
5. Hochreiter S, Clevert DA, Obermayer K (2006) A new summarization method for affymetrix probe level data. Bioinformatics 22(8):943–949
6. Laiho A, Elo LL (2014) A note on an exon-based strategy to identify differentially expressed genes in RNA-Seq experiments. PLoS One 9(12), e115964
7. Ohtaki M, Otani K, Hiyama K, Kamei N, Satoh K, Hiyama E (2010) A robust method for estimating gene expression states using affymetrix microarray probe level data. BMC Bioinformatics 11(1):183
8. Liu X, Milo M, Lawrence ND, Rattray M (2005) A tractable probabilistic model for affymetrix probe-level analysis across multiple chips. Bioinformatics 21(18):3637–3644
9. Liu X, Milo M, Lawrence ND, Rattray M (2006) Probe-level measurement error improves accuracy in detecting differential gene expression. Bioinformatics 22(17):2107–2113
10. Lemon WJ, Liyanarachchi S, You M (2003) A high performance test of differential gene expression for oligonucleotide arrays. Genome Biol 4(10):R67. doi:10.1186/gb-2003-4-10-r67
11. Master SR, Stoddard AJ, Bailey LC, Pan TC, Dugan KD, La C (2005) Genomic analysis of early murine mammary gland development using novel probe-level algorithms. Genome Biol 6(2):R20. doi:10.1186/gb-2005-6-2-r20
12. Lahti L, Torrente A, Elo LL, Brazma A, Rung J (2013) A fully scalable online preprocessing algorithm for short oligonucleotide microarray atlases. Nucleic Acids Res 41(10). doi:10.1093/nar/gkt229, 1212.5932v2
13. Abramowitz M, Stegun IA (1972) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol 55. Courier Dover Publications, New York
14. Laajala E, Aittokallio T, Lahesmaa R, Elo LL (2009) Probe-level estimation improves the detection of differential splicing in Affymetrix exon array studies. Genome Biol 10(7):R77
15. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40(12):1413–1415. doi:10.1038/ng0609-762d
16. Ta C, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, Williams A, Blume JE (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol 8(4):R64. doi:10.1186/gb-2007-8-4-r64
17. Cline MS, Blume J, Cawley S, Clark Ta HJS, Lu G, Salomonis N, Wang H, Williams A

(2005) ANOSVA: a statistical method for detecting splice variation from expression data. Bioinformatics 21(suppl 1):107–115. doi:10.1093/bioinformatics/bti1010

18. Affymetrix (2005) Alternative transcript analysis methods for exon arrays. http://www.affymetrix.com/support/technical/whitepapers.affx or techical notes
19. Bengtsson H, Simpson K, Bullard J, Hansen K (2008) aroma.affymetrix: a generic framework in R for analyzing small to very large affymetrix data sets in bounded memory. Tech. rep., Department of Statistics, University of California
20. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol 57:289–300
21. Affymetrix (2007) Identifying and validating alternative splicing events. http://www.affymetrix.com/support/technical/technotes-main.affx

Chapter 12

Profiling Changes in Histone Post-translational Modifications by Top-Down Mass Spectrometry

Mowei Zhou, Si Wu, David L. Stenoien, Zhaorui Zhang, Lanelle Connolly, Michael Freitag, and Ljiljana Paša-Tolić

Abstract

Top-down mass spectrometry is a valuable tool for understanding gene expression through characterization of combinatorial histone post-translational modifications (i.e., histone code). In this protocol, we describe a top-down workflow that employs liquid chromatography (LC) coupled to mass spectrometry (MS), for fast global profiling of changes in histone proteoforms, and apply LCMS top-down approach for comparative analysis of a wild-type and a mutant fungal species. The proteoforms exhibiting differential abundances can be subjected to further targeted studies by other MS or orthogonal (e.g., biochemical) assays. This method can be generally adapted for screening of changes in histone modifications between samples such as wild type *vs.* mutant or healthy *vs.* diseased.

Key words Histone, Post-translational modification, Liquid chromatography, Mass spectrometry, Top-down, Screening

1 Introduction

Revealing the complexity of protein species or proteoforms arising from sequence differences, proteolysis, and post-translational modifications (PTMs) is essential for understanding protein function and critically needed for identification of novel biomarkers [1]. Mass spectrometry (MS) has become an indispensable tool for analysis of proteoforms due to its ability to efficiently perform large-scale analysis of complex mixtures [2, 3]. The peptide level or "bottom-up" workflow is routinely used for characterization of protein primary sequence and identification of proteins extracted from biological samples. This process involves enzymatic digestion of proteins into short peptides. Liquid chromatography (LC) is typically introduced prior to MS to reduce sample complexity and facilitate more effective characterization of peptides. The peptide mass measurement, together with fragmentation of the peptides within the mass spectrometer, allows for the peptide sequence

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_12,

identification. The identified peptides are used as proxy for the corresponding proteins. With recent advances in instrument development, the "top-down" workflow can now also be applied for characterizing the whole intact protein (i.e., bypassing enzymatic digestion), instead of analyzing small peptides [2, 3]. Top-down MS preserves information about combinatorial PTMs on sequence variants and truncated forms of proteins and thus offers great opportunities for direct characterization of functional proteoforms present in the sample.

Histones are highly conserved, yet heavily modified eukaryotic proteins that are involved in gene expression and regulation. The histone PTMs are highly complex and are known to act as "codes" for silencing or activating genes [4–6]. Changes in one or a combination of multiple histone PTMs could trigger significant changes in gene expression and subsequently relate to certain diseases including cancer [7–9]. Various MS methods, including bottom-up, top-down, and middle-down (e.g., using less frequent cutters, such as Glu-C or Asp-N to generate longer peptides harboring multiple PTMs), have all been successfully utilized to characterize histone PTMs [10, 11]. Each method has its own advantages and resulting data can be highly complementary. In this protocol, we present a top-down LCMS workflow for fast screening of changes in core histone profiles using wild-type (WT) and a *kmt6* mutant of *Fusarium graminearum* as an example. The identified differences in major H3 proteoforms are consistent with results previously obtained by chromatin immunoprecipitation coupled to DNA sequencing (ChIP-seq) and Western blot analyses [12]. This workflow can be universally applied for fast global profiling of histone proteoforms and could lead to discovery of potential new targets for other biochemical and related studies. For instance, epigenetic alterations can be reversed making these an attractive target for cancer therapies.

2 Materials

2.1 Histone Extraction from Cells

1. HeLa core histones (catalog number 53501) were purchased from Active Motif (Carlsbad, CA) as a standard for optimizing LCMS experimental conditions.
2. Histone preparation solutions for *Fusarium graminearum*: buffer A (1 M sorbitol, 7% Ficoll type 400, 20% glycerol, 5 mM magnesium acetate, 3 mM $CaCl_2$, 5 mM ethylene glycol tetraacetic acid, 50 mM Tris–HCl, pH 7.5), buffer B (10% glycerol, 5 mM magnesium acetate, 5 mM ethylene glycol tetraacetic acid, 25 mM Tris–HCl, pH 7.5), and buffer C (1 M sucrose, 10% glycerol, 25 mM Tris–HCl, pH 7.5). Store all buffers at 4 °C and add 0.5 mM dithiothreitol (DTT) and protease inhibitors prior to use.

2.2 First Dimension Chromatography for Purification and Fractionation of Individual Histone Superfamilies

1. A custom in-house built LC, the design of which is described previously [13, 14], was used for off-line fractionation. Briefly, separation was carried out at 4000 psi with two syringe pumps (Model 100 DM, ISCO, Lincoln, NE). The gradient was generated by replacing mobile phase A1 with mobile phase B1 in a 2.5 mL stirred mixer (started with 100% mobile phase A1). A split line was used to obtain a flow rate around 20 μL/min from the pump and about 1 μL/min to the LC column.
2. In-house packed reversed-phase LC (RPLC) C5 column: Phenomenex (Torrance, CA, USA) Jupiter C5 5 μm particle, pore size 300 Å, column inner diameter (ID) 200 μm, outer diameter (OD) 360 μm, length 90 cm.
3. Mobile phase A1 composition (percent volume): 20% acetonitrile (ACN), 5% isopropyl alcohol (IPA), 0.6% formic acid (FA), and 74.4% water.
4. Mobile phase B1 composition (percent volume): 45% ACN, 45% IPA, 0.6% FA, and 9.4% water.
5. A UV detector was connected at the end of the LC column and before the fraction collector to monitor the elution of histones at the wavelength of 214 nm.
6. A Triversa NanoMate 100 (Advion Biosciences, Ithaca, NY) was used for automatic fraction collection onto a 96-well plate (Eppendorf, Westbury, NY).

2.3 Second Dimension Online Chromatography Coupled to Mass Spectrometry

1. A custom-built nanoLC system with Agilent nano-pumps controlled by LCMSNet [15] or a commercial Waters NanoAquity LC (two pumps) was used for the second dimension separation.
2. RPLC separation:
   (a) In-house packed RPLC C18 column: Phenomenex C18 3 μm particle, pore size 300 Å, column ID 75 μm, OD 360 μm, length about 50 cm.
   (b) Online solid phase extraction (SPE) trap column: Phenomenex Aeris WP C18 3.6 μm particle, column ID 150 μm, OD 360 μm, length about 10 cm.
   (c) Mobile phase A2 composition (percent volume), 1% FA in water; mobile phase B2 composition (percent volume), 1% FA in ACN.
3. Weak cation exchange-hydrophilic interaction LC (WCX-HILIC) separation:
   (a) In-house packed WCX-HILIC column: polyCAT A 5 μm particle, pore size 1000 Å, column ID 100 μm, OD 360 μm, length about 70 cm.
   (b) Online SPE trap column: polyCAT A 5 μm particle, pore size 1000 Å, column ID 150 μm, OD 360 μm, length about 10 cm.

(c) Mobile phase A3 composition (percent volume), 70% ACN, 1% FA in water; mobile phase B3 composition (percent volume), 70% ACN, 10% FA in water (*see* **Note 1**: low pH may etch the pumps).

2.4 Mass Spectrometers

The LTQ Orbitrap Velos mass spectrometer or the Velos Orbitrap Elite mass spectrometer with electron transfer dissociation (ETD) was used to acquire high-resolution MS^1 and MS^2 spectra.

2.5 Data Analysis Software

We have used in-house developed software (MSPathFinder, LcMsSpectator, and DeconTools), which can be downloaded from the PNNL Omics website at http://omics.pnl.gov/software.

3 Methods

3.1 Nuclei Isolation from Fungus

1. Isolate nuclei from 5 to 6 g of *Fusarium graminearum* mycelia previously stored at −80 °C. Grind mycelia into a fine powder using a mortar and pestle under liquid nitrogen for 5 min. Transfer powder and the remaining liquid nitrogen to loosely capped 50 mL conical tubes and allow liquid nitrogen to evaporate at −80 °C for ~30 min.
2. Add 15 mL of a 1:1.7 buffer A/buffer B mixture to four 30 mL Oakridge centrifuge tubes and store on ice.
3. Add 40 mL of ice-cold buffer A to tubes containing ground mycelia and mix by stirring until homogenous. Transfer mixture to a 250 mL flask and slowly add 70 mL of ice-cold buffer B and mix by swirling.
4. Filter mixture through two layers of cheese cloth. Carefully add 25 mL of the filtered cell extract to the top of the four previously prepared centrifuge tubes (in **step 2**).
5. Centrifuge at 2600 × *g* for 7 min in a swinging bucket rotor. Collect the top 20 mL of supernatant and layer onto 5 mL of buffer C in four 50 mL conical tubes.
6. Centrifuge at 7500 × *g* for 16 min. Carefully decant the supernatant and store pellets at −80 °C. The presence of nuclei can be validated by examining the residual buffer above the pellet in a microscope.

3.2 Histone Extraction from Nuclei Pellets

1. Histones are prepared using previously published purification methods [16] (*see* **Note 2**, alternative extraction method). Nuclei pellets are resuspended in 500 μl of 0.2 M H_2SO_4 by pipetting up and down until no clumps are visible. If necessary vortex the solution to break up clumps. Incubate for 30 min to overnight with rotation at 4 °C.

2. Nuclear debris is removed by centrifugation at 16,000 × *g* for 10 min at 4 °C. Remove the supernatant to a clean tube and precipitate histones by adding 1/3 volume of trichloroacetic acid one drop at a time. Invert the tube several times to thoroughly mix and incubate the tube on ice for 30 min to overnight. The solution will become slightly cloudy over time as the histones precipitate.
3. Pellet histones by centrifugation at 16,000 × *g* for 10 min at 4 °C. Remove supernatant with pipette taking care not to disturb the pellet which may be on the side wall of the tube.
4. Wash the pellet to remove acid by adding 1 mL of ice-cold acetone. Centrifuge at 16,000 × *g* for 5 min and then carefully remove the supernatant. Repeat this acetone wash a total of three times. Allow the final pellet to air-dry for ~20 min at room temperature.
5. Dissolve the pellet containing histones in 100 μl of H_2O by pipetting up and down. Histone pellets are often smeared on the wall of the tube so the resuspension process should account for this. Remove the dissolved histones to a clean tube. If visible aggregates remain, these can be removed by centrifugation at 16,000 × *g* for 10 min at 4 °C. Store purified histones at −80 °C until use. If necessary aliquot histones to prevent repeated freeze-thawing.
6. Oxidize the extracted histones by incubating with 3% hydrogen peroxide and 3% formic acid at room temperature for 4 h [17] (*see* **Note 3**: intentional oxidation of sample if needed). Proceed to the next step immediately to quench the reaction.

3.3 First Dimension Chromatography Purification and Fractionation of Histone Sample

1. Inject extracted histones (about 10 μg) onto the C5 column. C5 RPLC separation helps quench the oxidation and improve purity of the sample. Alternatively, the histones can be fractionated for targeted analysis of each histone family (*see* **Note 4**: the pros and cons of fractionation).
2. Load the sample on the column with solvent A1 for 20 min (flush with at least twice the volume of the sample loop). Measure the flow rate of the solution eluting from the column and record it for calculating the dead time from the UV detector (transfer line length divided by flow rate).
3. Change the solvent flowing into the reservoir from solvent A1 to solvent B1. Gradient is started by gradually replacing the prefilled A1 solvent in the mixer reservoir with solvent B1. Start UV chromatogram and collection of the fractions into a 96-well plate using the automated TriVersa NanoMate system (*see* **Note 5**: add water to the wells). Histones generally elute around 100–150 min for the LC described earlier. Use the distinct profile of HeLa histone standard as retention time reference for your own LC system.

4. Manually combine fractions based on the UV elution profile (*see* **Note 6**: choice of fractionation). Figure 1a shows a UV trace collected during the separation of HeLa core histone (a standard model used to optimize system performance). A representative UV chromatogram obtained for the wild-type *Fusarium graminearum* histones is shown in Fig. 1b (*see* **Note 7**: elution profile of histones can greatly vary between different organisms/samples). The fractions from the highlighted region in Fig. 1b are combined for the global profiling described below. For targeted analysis of individual histone family, each major peak in UV chromatogram can be collected as a separate fraction.
5. Concentrate the combined fractions by SpeedVac for the second dimension LCMS.

3.4 Second Dimension Online LC Separation Coupled to Online Mass Spectrometry Detection

1. Inject 2 μg of purified/fractionated histones of interest onto the second dimension LC coupled directly to MS (*see* **Note 8**: load amount). Use C18 for the second dimension LC separation of recombined fractions (i.e., all core histones) or for targeted analysis of fractionated H2A, H2B, and H3 families. Use WCX-HILIC as second dimension LC for targeted analysis of

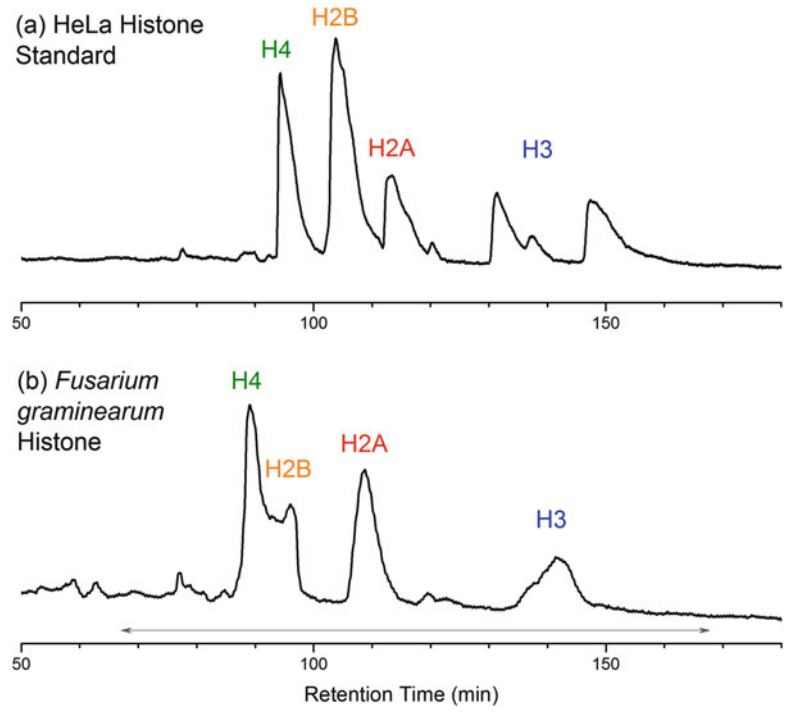

Fig. 1 Separation of (**a**) HeLa and (**b**) *Fusarium graminearum* core histones on C5 RPLC with UV detection. Note that different organisms/cell lines may show different chromatographic patterns. The *gray arrow* in (**b**) above the *horizontal axis* shows the retention time range where the fractions were combined for subsequent LCMS analysis

H4 proteoforms. Representative LCMS total ion chromatograms of H4, H2B, and H2A fractions with RPLC or WCX-HILIC columns are shown in Fig. 2.

2. Load the injected samples onto the SPE trap column using solvent A2 for RPLC or A3 for WCX-HILIC (i.e., 15–20 min).
3. Switch the eluent flow from the SPE column to the analytical column of choice. Start the gradient and mass spectrometry acquisition. In RPLC, histone proteins generally elute between 30 and 50% solvent B2 with the C18 column and conditions specified. For profiling all histone families, use a long gradient (i.e., 400 min) to maximize coverage.
4. Acquire MS^1 parent ion mass spectra at resolution of 120 K (a predefined factory setting for the mass analyzer where 120 K resolution can be achieved at *m/z* of 400). Use high micro-scan values (e.g., 7 or 8) (*see* **Note 9**: benefit of more MS^1 micro-scans). AGC targets are 5E6 for MS^1 and 5E5 for MS^2. Maximum fill time is 800 ms for MS^1 and 700 ms for MS^2.
5. Acquire data-dependent MS^2 fragment ion mass spectra for top six most abundant species with ETD, at resolution of 120 K (at *m/z* of 400) with dynamic exclusion and micro-scan of 4–6 (*see* **Note 10**: recommendations for MS^2 settings). ETD reaction time is set to 15–20 ms. Figure 3 shows a representative fragment ion coverage map for ETD and HCD of human H2A histone. Typically, ETD yields extensive N(C) terminal fragmentation resulting in better coverage and ability to confidently identify and locate PTMs on histones (*see* **Note 11**: ETD vs. HCD).

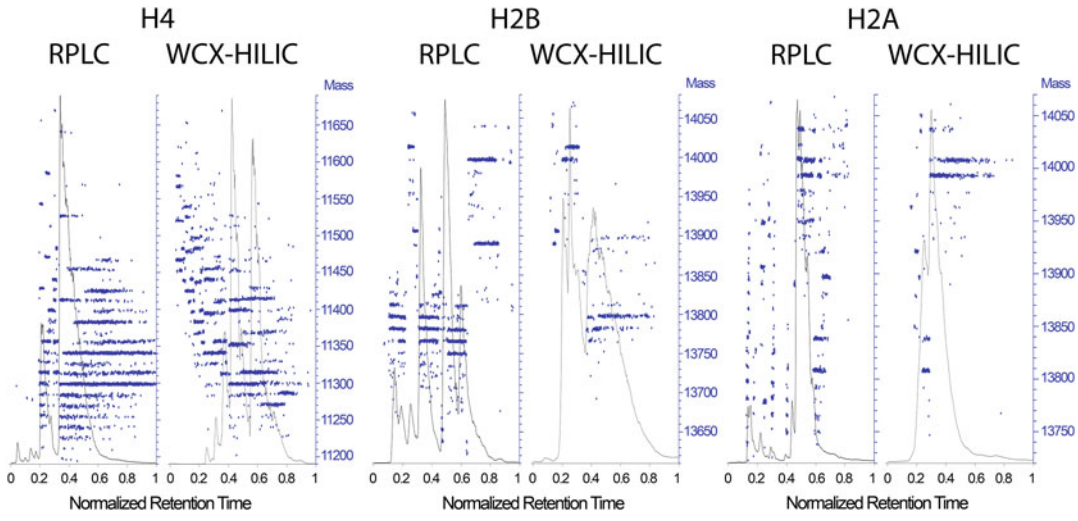

Fig. 2 Various LC separation modalities can be applied to further separate fractionated HeLa histone H4, H2B, and H2A families. The total ion chromatograms are overlaid with the deconvoluted masses shown by the blue dots (*blue vertical axis* on the *right* for mass). WCX-HILIC is best suited for H4 histones differing in the degree of acetylation, while H2A and H2B differing in primary amino acid sequence are more effectively separated by C18 RPLC

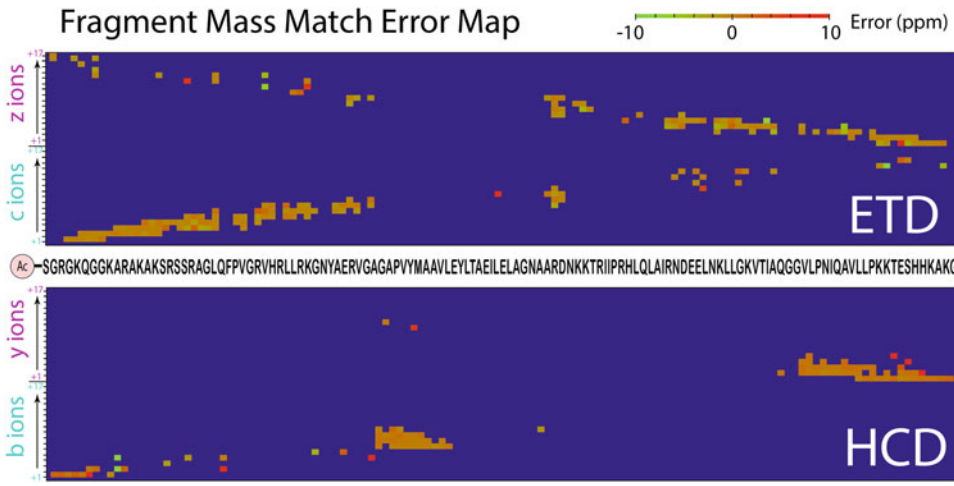

Fig. 3 Fragment error map for ETD (25 ms) and HCD (28 V) spectra acquired for human H2A type 2-A, generated using LcMsSpectator. The mass error is color coded (scale in the *top right corner*). Fragment ions are plotted along the protein sequence (shown in the *middle*) with c_1/b_1 starting on the left and z_1/y_1 starting on the *right*. Different charge states of a particular fragment ion are plotted along the *vertical axis*. ETD is typically more effective in generating good sequence coverage at the two termini and more effective in locating the PTM sites. HCD/CID typically show preferential cleavage at certain residues and can sometimes provide information complementary to ETD

3.5 Data Analysis

1. Search the datasets using MSPathFinder (*see* **Note 12**: brief software description) against all the histone sequences for *Fusarium graminearum* with internal cleavage and commonly expected modifications: Lys mono-/di-/trimethylation, Lys acetylation, Ser/Thr/Tyr phosphorylation, Met mono-/di-oxidation, and Cys tri-oxidation (*see* **Note 13**: choice of search parameters). (Detection of a residual abundant proteins, e.g., ribosomal proteins, is almost unavoidable.) Mass error tolerance is 10 ppm. MS^2 spectra can also be analyzed by MS-Align+ [18] (http://bix.ucsd.edu/projects/msalign/) and/or ProSightPC [19] (Thermo Fisher Scientific). Each tool has its own pros and cons based on specific applications (*see* **Notes 14** and **15** for description of these additional tools). Cross-reference between the outputs from multiple tools is recommended for more reliable results.
2. A "feature map" showing the mass of the detected species plotted against LC retention time can be visualized in LcMsSpectator simply by loading the output from MSPathFinder. An example showing the wild-type *Fusarium graminearum* histone data is given in Fig. 4.
3. We are using the difference in the normalized abundances of the LCMS features to determine and visualize the changes in

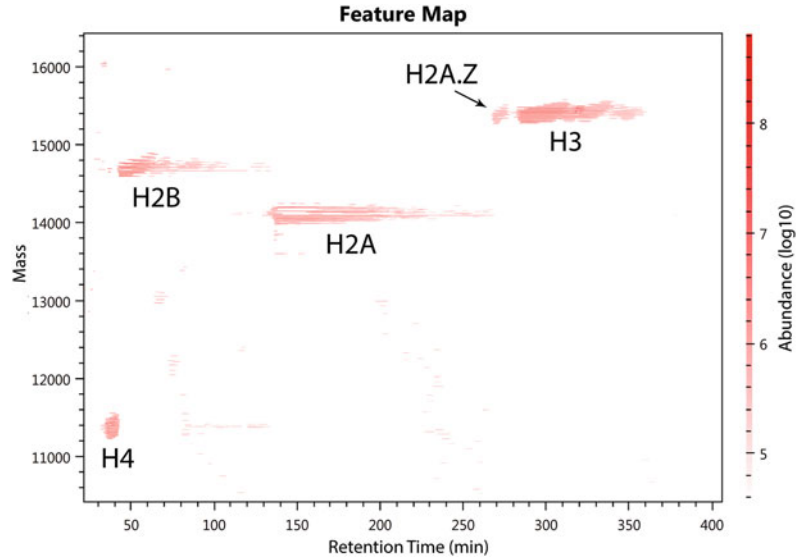

Fig. 4 A representative feature map generated from LCMS analysis of histones isolated from wild-type *Fusarium graminearum*, visualized in LcMsSpectator. Major histone species can be identified based on their retention times and accurate masses

histone proteoforms between the samples (e.g., mutant vs. wild-type) (*see* **Note 16**: data processing details). For *Fusarium* example, the differential abundances are plotted in Fig. 5 (*see* **Note 17**: finding LCMS features can be challenging). LCMS features displaying differential abundance are selected as targets for further analysis (*see* **Note 18**).

4. Search the list of identified proteoforms and locate the spectra showing significant differences in the feature map. Manually check parent (MS^1) and fragment ion (MS^2) spectra to validate the identification. For instance, the H3 species at 15,383 Da eluting around 285 min in the wild-type and species at 15,341 Da eluting around 292 min in the mutant were selected as interesting targets displaying the most abundant H3 features that are differentially abundant between the two samples. The MS^2 fragment ion spectra within the region of interest confirmed these proteoforms are differentially modified, in agreement with previous ChIP-seq results (Fig. 6, *see* **Note 19**). For a good match, most of the major fragment ions should be accounted for, display similar mass accuracy, and the precursor mass should be within the isolation window (Fig. 7, *see* **Note 20**).
5. Differentially abundant proteoforms between wild-type and mutant identified in the global profiling can be used as targets

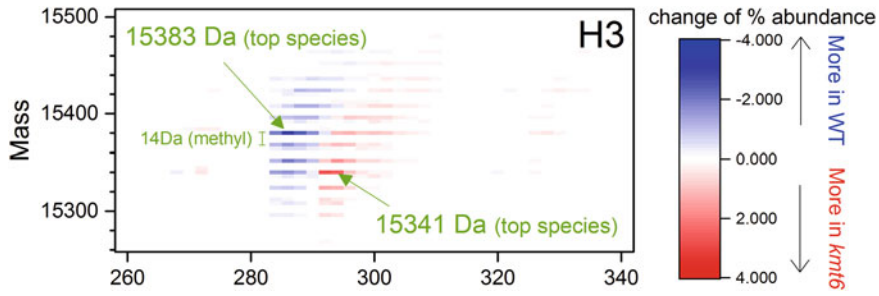

Fig. 5 Differential map of H3 histones (% abundance in *kmt6* vs. wild-type). LCMS features shown in *blue points* are more abundant in wild-type (WT), and LCMS features shown in *red* are more abundant in *kmt6* mutant. Abundance was normalized to the total ion signal for all detected H3 histones. The most abundant H3 species in the WT has a mass about 42 Da higher and a retention time slightly smaller than the most abundant species in the mutant. Their MS^2 spectra were manually examined for further verification. The x-axis denotes the retention time in minutes

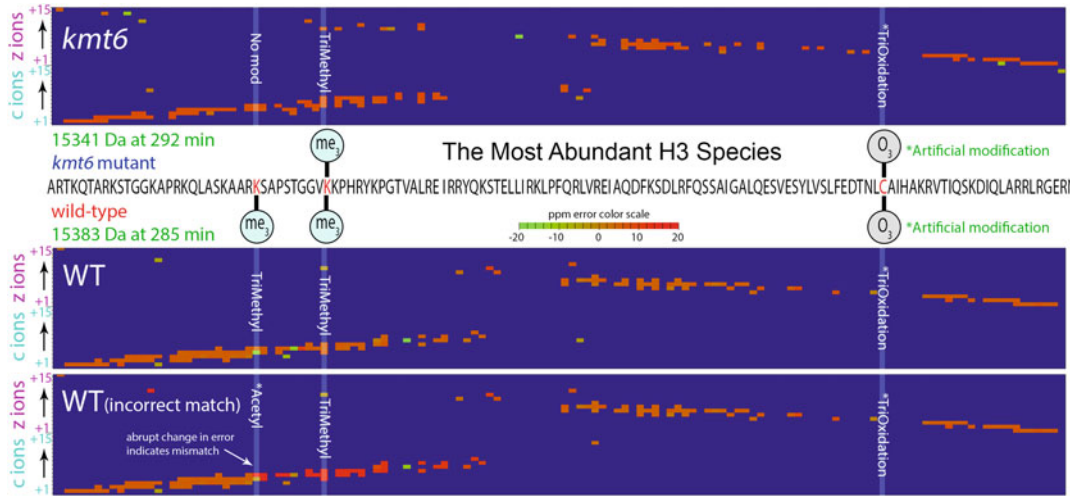

Fig. 6 Electron transfer dissociation (ETD) error maps for the identified top H3 species in the wild-type and the *kmt6* mutant. Fragment mass errors are plotted along the sequence using the color scheme in the legend shown in the *middle*. High-resolution and mass accuracy tandem mass spectra facilitate differentiation between acetylated and trimethylated proteoforms ($\Delta m = 36$ mDa)

for further validation studies and guide the selection of specific proteoforms to better understand the gene control mechanism.

4 Notes

1. The low pH mobile phase can etch LC parts; thus, it is recommended to flush the LC pumps with regular solvents for long time storage after using 10% formic acid as mobile phase.

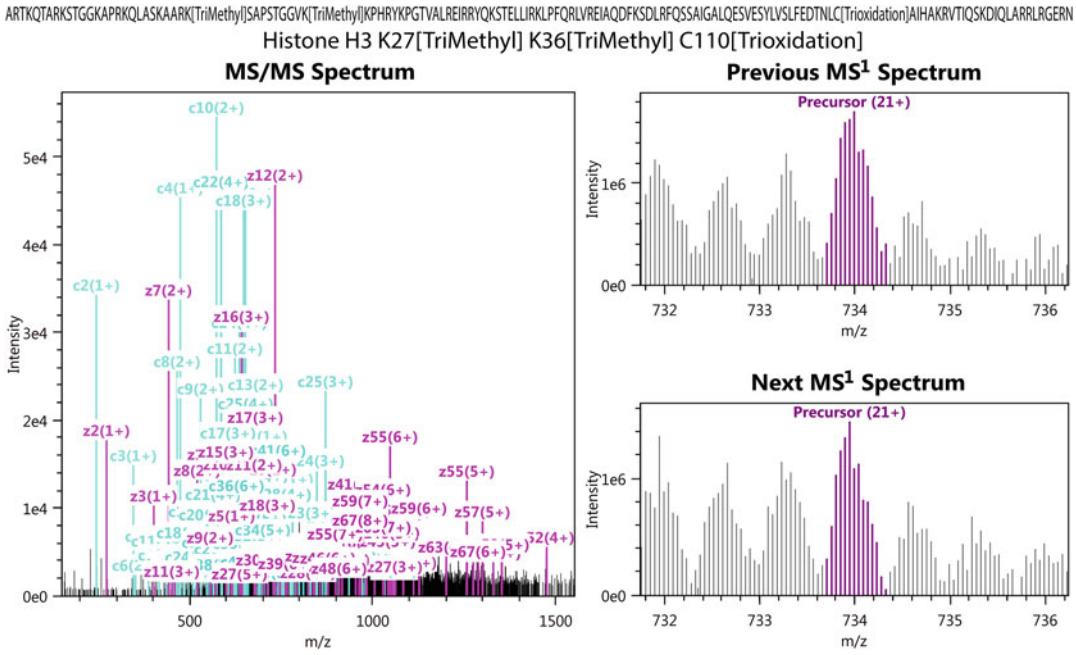

Fig. 7 Illustration of a highly confident identification using LcMsSpectator for manual examination (i.e., the most abundant H3 species at 15,383 Da in WT). The major fragments are matched in the tandem mass spectrum, and the precursor ion matches the expected proteoform in adjacent (i.e., previous and the following) parent ion spectra within the isolation window of 3 Th around $m/z = 734$ (x axis centered at the isolated precursor m/z)

2. The nuclei extraction protocol described herein was optimized for *Fusarium graminearum*. Alternative methods to purify nuclei may also be used, particularly for dissimilar organisms. For instance, when working with cell lines, whole cell extracts may be used. In addition, we have used a proprietary histone purification kit (i.e., Active Motif) to effectively purify histones from mammalian cell lines. However, this approach appears to be ineffective for purifying *Fusarium graminearum* histones.
3. The histone extraction protocol that employs sulfuric acid or other oxidizing reagents can generate significant and highly variable amount of oxidized histones (cysteine and methionine). Incubating the histone samples with 3% formic acid and 3% hydrogen peroxide at room temperature for 4 h will fully oxidize cysteine and methionine residues [17] and is recommended in order to avoid misinterpretation. Namely, artificial oxidation can be very problematic when comparing histone proteoforms between samples because some detected differences may come from a different extent of oxidation during sample preparation rather than underlying biology. If oxidation is a biologically relevant targeted PTM, other extraction methods should be used and the oxidation step should be skipped.

4. Fractionating histones into individual chains (H4, H2B, H2A, H3) and analyzing them separately by LCMS can improve the overall sensitivity and coverage. This step helps remove some other proteins that are carried over during the extraction step. In addition, optimum separation conditions can be chosen for each histone family when they are fractionated into different samples (e.g., H4 can be fractionated from other histone families by C5 RPLC and subject to WCX-HILIC for optimum separation of proteoforms as shown in Fig. 2). However, the fractionation step will significantly increase the number of the LCMS analyses and thus reduce throughput. Whether to skip the fractionation or not will depend on the application and the purity of the sample. Clearly, for targeted analysis of individual histone family, the fractionation step can be very useful.
5. Add 10–20 μL of water into the wells in the collection plate, depending on the volume of each fraction. This is to prohibit complete evaporation of the liquid for low-volume fractions (typically 2 μL per fraction for the fractionation step in this protocol).
6. Typically, the fractionation step is not necessary for global screening across all core histones. Simply combine the core histone fractions and use as purified sample for LCMS.
7. Throughout this work, we are using HeLa core histone sample as a standard because it shows well-resolved peaks for each histone family under the experimental conditions specified. It can be used as a standard for optimizing the experimental conditions. However, different samples can have very different elution profiles depending on the sample complexity and host organism (such as the case for the histones isolated from *Fusarium graminearum*).
8. Optimum sample load depends on the size of the column used and also the sensitivity of the instrument. It is important not to overload and alternate blank between sample analyses when performing comparative studies to minimize the carryover. For comparative studies, it is important to inject similar amount of proteins for each analysis. The actual concentrations of the histone proteins can be estimated from the UV chromatogram during the fractionation step. It is not uncommon that other proteins are co-purified with histones, and the protein concentration derived using bulk assays (e.g., bicinchoninic acid assay) may overestimate the actual histone concentration.
9. Larger numbers of micro-scans can significantly improve the quality of spectra, especially for the H3 family, which normally displays much lower S/N than the other core histone families. Better quality spectra will allow higher confidence in proteoform identification/characterization.

10. Fragment ion spectra (MS^2) should be acquired for as many unique proteoforms as possible to maximize the number of identifications. Dynamic exclusion is often used to reduce the chance of fragmenting the same species over and over again by creating an exclusion list of ions (m/z) that have been fragmented within a certain time period during acquisition. However, many different histone proteoforms will have the exact same mass, so it is not ideal to set a long duration for dynamic exclusion. Better strategy is to specify narrow m/z range for selecting parent ions (e.g., $700 < m/z < 900$) where the most abundant histone ions are typically detected to prevent wasteful fragmentation of multiple charge states of the same (highly abundant) species.
11. ETD generally yields very good sequence coverage at the N-terminus and C-terminus of the protein. Most of the histone PTMs are located near the N-terminus, making ETD a method of choice for histone characterization. CID/HCD tend to give sequence specific fragments, which can sometimes be useful in providing complementary information because ETD typically does not cover the middle of the protein sequence. However, in most cases, ETD is more effective in locating the PTM sites.
12. MSPathFinder is an in-house developed search engine, with accessory data viewer LCMSSpectator for manual data validation. A FASTA database including all histone sequences is required. All possible modifications need to be defined in the search. Otherwise it will generate false identifications for proteoforms with unexpected modifications.
13. MSPathFinder is an efficient strategy for targeted searches. Searching with single cleavage is recommended because it is common to see fragments in histone samples. However, searching a large protein database with single cleavage and large number of PTMs can be computationally prohibitive. It is therefore recommended to perform a search with MS-Align+ against a large database to explore if there are any unexpected modifications or other proteins that should be included in the MSPathFinder search. When correct PTM or protein is not included in the database, MSPathFinder tends to force the match to a combination of truncations and specified PTMs, or it may not identify any proteoform at all. Such false positives can be easily cross-checked and manually fixed with information from MS-Align+.
14. MS-Align+ is an open modification search engine generally applicable to top-down MS data analysis. Only a FASTA database including the target proteins is required. It reports the mass shift values for modifications on identified proteins and the range of residues that could be modified.

15. ProSightPC is another useful tool for characterization of histone PTMs, commercially available from Thermo Fisher Scientific. The tool is unique in that it uses pre-annotated database, which can be generated by downloading the text format of the sequences from UniProt database. In general, it works well for organisms with well-annotated protein sequences (e.g., human histones). However, for poorly studied/annotated organisms, the applicability of ProSightPC could be quite limited.
16. The data files are deconvoluted with DeconTools, and the output peak lists are binned into matrices with bin sizes of 2 min in retention time and 4 Da in mass. The binning step helps "smooth" out minor variations for better visualization of major differences in the LCMS feature maps between the samples.
17. It can be sometimes difficult to tell if the differential LCMS features originate from the drift in LC retention time or if they reflect real differences in PTMs. Ideally, multiple LCMS feature maps from replicate analyses should be acquired, aligned, and normalized for statistically sound comparison. However, several histone proteoforms display very similar (e.g., trimethylation vs. acetylation corresponds to 36 mDa difference in mass) or exactly the same (e.g., proteoforms differing in the PTM site) mass that cannot be resolved in MS. In some cases these proteoforms may be separated by LC retention time but in most cases the samples are too complex to be resolved using conventional LCMS. Standard alignment algorithms typically "over-align" the histone LCMS data because large number of proteoforms with similar/identical mass does not offer a good reference point for alignment by accurate mass. Using optimized LC conditions and analyzing the samples close in time will minimize the drift in retention and is recommended for better alignment between datasets. Development for new alignment algorithms for histone LCMS features is work in progress.
18. It is important to precisely quantify the differences in histone proteoforms' relative abundances. Even a complete loss of a particular proteoform may not generate a complete elimination of the corresponding LCMS feature, because there can be other isoforms not fully resolved at the same (or similar) retention time (e.g., deletion of H3K27me3 in the mutant does not eliminate the presence of 15,383 Da because proteoforms displaying trimethylation or acetylation at other positions elute at similar retention times; however, the overall abundance of the 15,383 Da species will change). Thus, the change in relative abundance of LCMS features is a more sensitive measure of changes in histone proteoforms. The fungal species used in this study display a relatively simple histone profile making the

manual analysis relatively straightforward. However, histone profiles from human and other organisms can be very complex and difficult to analyze manually (esp. H3 proteoforms). With good experimental design and adequate number of biological and technical replicates, the quantitative differences can be determined more accurately. The software for determining and reporting differential LCMS features is currently under development. An alternative way is to integrate with middle-down and/or bottom-up data to obtain different levels of information that suit specific goals of the study [20].

19. The error map allows for quick check of the quality of the match. A high sequence coverage is an indication of high confidence in the PTM assignments. Additionally, high mass resolving power and accuracy instruments (e.g., FTMS) facilitate confident differentiation between acetylation (42,010 Da) and trimethylation (42,047 Da). Using typical search algorithms, it is often challenging to differentiate between these two PTMs because both will match the data within typical mass error tolerance (e.g., 10 ppm). However, with high enough mass measurement accuracy, a false match will normally display an abrupt change in the fragment mass error (e.g., as indicated by an abrupt change in color in the fragment error map in Fig. 6).
20. Co-eluting proteoforms is a common occurrence in histone top-down analysis. Therefore, we are often dealing with multiplexed tandem mass spectra composed of fragments from multiple, co-eluting proteoforms that cannot be resolved in LC. For instance, H3 proteoforms differing in degree of methylation (+1/−1) usually co-elute and are often co-selected (with typical isolation window of 3 Th). The proteoform that can be mapped with the largest number of high-intensity fragment ions is at present the most plausible identification. Hence, significant developments in instrumentation (separations and MS) and bioinformatics are needed to advance this field in relation to epigenetic research.

Acknowledgment

The authors thank Christopher Wilkins, Jung Kap Park, and Sangtae Kim at the Pacific Northwest National Laboratory (PNNL) for developing the bioinformatics software used in this work. We appreciate the help from other PNNL colleagues: Matthew Monroe and Nikola Tolić for data analysis and Rosalie K. Chu, Rui Zhao, Anil K. Shukla, and Ron Moore for running LCMS experiments. We also thank Jonathan Galazka at the Oregon State University for preparing the fungal histone samples. The research was performed

in Environmental Molecular Sciences Laboratory (EMSL), a US Department of Energy (DOE) national user facility at the Pacific Northwest National Laboratory (PNNL) in Richland, WA.

References

1. Smith LM, Kelleher NL (2013) Proteoform: a single term describing protein complexity. Nat Methods 10(3):186–187. doi:10.1038/nmeth.2369
2. Han X, Aslanian A, Yates Iii JR (2008) Mass spectrometry for proteomics. Curr Opin Chem Biol 12(5):483–490. doi:10.1016/j.cbpa.2008.07.024
3. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 11(1):49–79. doi:10.1146/annurev-bioeng-061008-124934
4. Bowman GD, Poirier MG (2015) Post-translational modifications of histones that influence nucleosome dynamics. Chem Rev 115(6):2274–2295. doi:10.1021/cr500350x
5. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403(6765):41–45
6. Jenuwein T, Allis CD (2001) Translating the histone code. Science 293(5532):1074–1080. doi:10.1126/science.1063127
7. Chi P, Allis CD, Wang GG (2010) Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 10(7):457–469. doi:10.1038/nrc2876
8. Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13(5):343–357. doi:10.1038/nrg3173
9. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28(10):1057–1068. doi:10.1038/nbt.1685
10. Yuan Z-F, Arnaudo AM, Garcia BA (2014) Mass spectrometric analysis of histone proteoforms. Annu Rev Anal Chem 7(1):113–128.doi:10.1146/annurev-anchem-071213-015959
11. Moradian A, Kalli A, Sweredoski MJ, Hess S (2014) The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. Proteomics 14(4-5):489–497. doi:10.1002/pmic.201300256
12. Connolly LR, Smith KM, Freitag M (2013) The *Fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLoS Genet 9(10), e1003916. doi:10.1371/journal.pgen.1003916
13. Tian Z, Tolic N, Zhao R, Moore R, Hengel S, Robinson E, Stenoien D, Wu S, Smith R, Pasa-Tolic L (2012) Enhanced top-down characterization of histone post-translational modifications. Genome Biol 13(10):R86. doi:10.1186/gb-2012-13-10-r86
14. Wu S, Lourette NM, Tolić N, Zhao R, Robinson EW, Tolmachev AV, Smith RD, Paša-Tolić L (2009) An integrated top-down and bottom-up strategy for broadly characterizing protein isoforms and modifications. J Proteome Res 8(3):1347–1357. doi:10.1021/pr800720d
15. LaMarche BL, Orton DJ, Clark DA, Ryan JD, Hopkins DF, Anderson GA, Moore RJ, Smith RD LC control software for maximum flexibility and rapid automation of new mode of operation. In: 59rd ASMS conference on mass spectrometry and allied topics, Denvor, CO, 2011
16. Shechter D, Dormann HL, Allis CD, Hake SB (2007) Extraction, purification and analysis of histones. Nat Protoc 2(6):1445–1457
17. Pesavento JJ, Garcia BA, Streeky JA, Kelleher NL, Mizzen CA (2007) Mild performic acid oxidation enhances chromatographic and top down mass spectrometric analyses of histones. Mol Cell Proteomics 6(9):1510–1526. doi:10.1074/mcp.M600404-MCP200
18. Liu X, Sirotkin Y, Shen Y, Anderson G, Tsai YS, Ting YS, Goodlett DR, Smith RD, Bafna V, Pevzner PA (2012) Protein identification using top-down spectra. Mol Cell Proteomics 11(6). doi:10.1074/mcp.M111.008524
19. LeDuc RD, Taylor GK, Kim Y-B, Januszyk TE, Bynum LH, Sola JV, Garavelli JS, Kelleher NL (2004) ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. Nucleic Acids Res 32(suppl 2):W340–W345. doi:10.1093/nar/gkh447
20. Sidoli S, Lin S, Karch KR, Garcia BA (2015) Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. Anal Chem 87(6):3129–3133. doi:10.1021/acs.analchem.5b00072

Chapter 13

Determining if an mRNA is a Substrate of Nonsense-Mediated mRNA Decay in *Saccharomyces cerevisiae*

Marcus J.O. Johansson

Abstract

Nonsense-mediated mRNA decay (NMD) is a conserved eukaryotic quality control mechanism which triggers decay of mRNAs harboring premature translation termination codons. In this chapter, I describe methods for monitoring the influence of NMD on mRNA abundance and decay rates in *Saccharomyces cerevisiae*. The descriptions include detailed methods for growing yeast cells, total RNA isolation, and Northern blotting. Although the chapter focuses on NMD, the methods can be easily adapted to assess the effect of other mRNA decay pathways.

Key words Yeast, mRNA levels, mRNA decay, Half-lives, RNA isolation, Northern blotting, NMD

1 Introduction

mRNA degradation is an important aspect of gene expression as it controls mRNA levels and eliminates aberrant transcripts. This chapter describes methods for determining if a transcript is a substrate of nonsense-mediated mRNA decay (NMD) in *S. cerevisiae*. NMD is an mRNA surveillance mechanism that triggers decay of transcripts harboring premature translation termination codons [1, 2]. The destabilization of such transcripts requires their translation and a distinct set of trans-acting factors, including the Upf1, Upf2, and Upf3 proteins [1, 2]. The substrates of NMD not only include transcripts from nonsense alleles but also other mRNAs with stop codons in a premature or atypical context. In yeast, the NMD substrates include intron-containing pre-mRNAs that are exported to the cytoplasm, transcripts from pseudo- and bicistronic genes, mRNAs in which ribosomes scan past the initiation codon and initiates at a downstream out-of-frame AUG, transcripts from transposable elements or their long terminal repeats, mRNAs encompassing unusually long 3′-UTRs, mRNAs containing a programmed frameshifting site, RNAs derived from transcriptional noise, and mRNAs containing upstream open reading frames (uORFs) [2–5].

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_13,

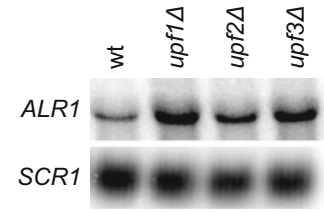

Fig. 1 Effect of NMD inactivation on *ALR1* mRNA abundance. Northern analysis of total RNA isolated from wild-type (MJY142), *upf1Δ* (MJY67), *upf2Δ* (MJY169), and *upf3Δ* (MJY150) cells grown in SC medium at 30 °C. The blot was probed for *ALR1* and *SCR1* transcripts using randomly labeled DNA fragments. The noncoding *SCR1* transcript serves as the loading control

Transcripts regulated by NMD usually accumulate in cells deleted for *UPF1*, *UPF2*, or *UPF3*. However, increased mRNA levels can arise either directly through a change in the decay rate or indirectly through the stabilization of mRNAs encoding regulatory proteins, which means that measurements of mRNA decay rates are essential when determining if a transcript is a *bona fide* NMD substrate. In addition to a protocol for determining mRNA abundance, I describe two different approaches for inhibiting transcription and monitoring mRNA decay. The first takes advantage of the regulatable *MET3* promoter (P_{MET3}) [6, 7], and the second utilizes a temperature-sensitive allele of a gene for a subunit in RNA polymerase II (*rpb1-1*) [8]. In the P_{MET3} system, the endogenous copy of the gene of interest is placed under control of P_{MET3}, and transcription is repressed by addition of methionine to the growth medium [6, 7]. Transcription inhibition in strains with an *rpb1-1* allele is achieved by shifting the cells from the permissive (25 °C) to the nonpermissive temperature (37 °C) [8, 9]. In all experiments, quantitative information about the RNA of interest is obtained by Northern blotting [10]. An advantage of Northern blotting is that it allows for an estimation of RNA size, which is, for example, important when the NMD substrate is an inefficiently spliced pre-mRNA. As examples, I show results from experiments where the abundance (Fig. 1) and decay rate (Fig. 2) of *ALR1* mRNA, a physiologically relevant NMD substrate [11], were determined in wild-type and NMD-deficient yeast cells. The *ALR1* gene encodes yeast cells' main magnesium importer [12, 13], and the *ALR1* mRNA is a NMD substrate due to the presence of a uORF(s) in the 5′-UTR [11].

2 Materials

All media and solutions are prepared using ultrapure water. Buffers and solutions for RNA work are prepared using RNase-free water and chemicals. Consult material safety data sheets and local regulations for appropriate handling of hazardous materials.

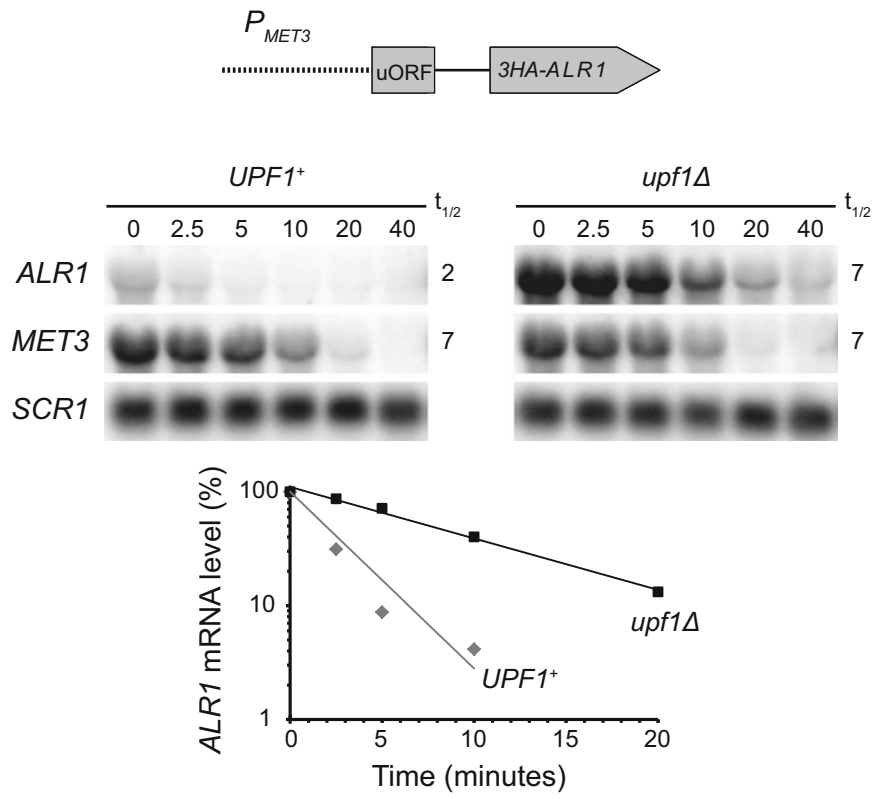

Fig. 2 Inactivation of NMD reduces the decay rate of *ALR1* mRNA. Northern analysis of total RNA isolated from P_{MET3}-*ALR1*$_{5'\text{-}UTR(-102)}$-*3HA-ALR1* (MJY484) to P_{MET3}-*ALR1*$_{5'\text{-}UTR(-102)}$-*3HA-ALR1 upf1Δ* (MJY485) cells following repression of P_{MET3} transcription by addition of methionine. Time points (minutes) after addition of methionine are indicated above the lanes. The blot was probed for *ALR1*, *MET3*, and *SCR1* transcripts using randomly labeled DNA fragments. The signal in each lane was quantified and normalized to the corresponding *SCR1* signal and the value expressed relative to the value for time point 0. The half-life ($t_{1/2}$, in minutes) was determined from the initial slope of the decay curve. The curves of *ALR1* mRNA in MJY484 (*UPF1*$^+$) and MJY485 (*upf1Δ*) are shown below

2.1 Yeast Strains and Media

To assess if a transcript is controlled by NMD, strains that are isogenic except for the allele at the *UPF1*, *UPF2*, or *UPF3* locus are used. Standard methods and genetic procedures are used to construct strains with the relevant alleles [14–16] (*see* **Note 1**). The strains used in the examples of this chapter (Figs. 1 and 2) are all congenic with MJY142 (*MATa ura3 leu2-2 his3-11,15 trp1 ade2-1 can1-100*) [11], including MJY67 (*upf1::HIS3MX6*), MJY169 (*upf2::kanMX6*), MJY150 (*upf3::kanMX6*), MJY484 (P_{ALR1}*::kanMX6-*P_{MET3}*-ALR1*$_{5'\text{-}UTR(-102)}$*-3HA*), and MJY485 (*PALR1::kanMX6-PMET3-ALR15′-UTR(-102)-3HA upf1::HIS3MX6*). In MJY484 and MJY485, the P_{MET3} sequence is integrated at the *ALR1* locus upstream of the sequence for the NMD-inducing uORF (Fig. 2) [11]. Yeast cells are grown in synthetic complete (SC) or SC-met-cys

medium [16], containing 0.67% yeast nitrogen base without amino acids, 2% glucose, and 0.2% of the appropriate dropout mix (*see* **Note 2**).

2.2 Reagents for RNA Preparation

Buffer A: 50 mM NaOAc, pH 5.2, 10 mM EDTA.

Complete buffer A: 45 mM NaOAc, pH 5.2, 9 mM EDTA, 1% SDS.

Phenol saturated with buffer A.

Phenol/chloroform/isoamyl alcohol (25:24:1 v/v), pH 6.6.

Ethanol (99.5% and 80%).

3 M NaOAc, pH 5.2.

2.3 Reagents and Equipment for Northern Blotting

10× MOPS electrophoresis buffer: 0.4 M MOPS, 100 mM NaOAc, 10 mM EDTA, pH 7.0.

Formaldehyde (36.5–38%).

RNA loading buffer: 50% deionized formamide, 1× MOPS buffer, 2.2 M formaldehyde.

3% glycerol, 0.02% bromophenol blue, 0.02% xylene cyanol, 25 μg/ml ethidium bromide (*see* **Note 3**).

20× SSC: 3 M NaCl, 300 mM sodium citrate, pH 7.0.

VacuGene XL vacuum blotting system.

Nylon blotting membrane, e.g., Zeta-Probe.

UV crosslinker, e.g., Stratalinker.

Pre-hybridization solution: 50% deionized formamide, 5× SSPE, 10× Denhardt's solution, 1% SDS, 100 μg/ml sonified salmon sperm DNA (*see* **Note 4**).

Hybridization solution: 50% deionized formamide, 5× SSPE, 2× Denhardt's solution, 5% dextran sulfate, 1% SDS, 100 μg/ml sonified salmon sperm DNA (*see* **Note 4**).

^{32}P-labeled random-primed DNA probes (*see* **Note 5**).

Wash buffer: 0.1× SSC, 0.1% SDS.

Phosphorimaging system, e.g., Typhoon FLA 9500.

3 Methods

3.1 Growth of Yeast Cells

Yeast cells are in all experiments grown in a shaking water bath at 150 rpm. For analyses of mRNA levels, an overnight culture is diluted to an optical density at 600 nm (OD_{600}) of ~0.025 in 20 ml of SC medium (*see* **Notes 6** and **7**), and the growth is monitored until the culture reaches OD_{600} ~ 0.5. Cells from 10 ml of the culture are harvested in a 15 ml tube by centrifugation at 1500 × *g* for 1 min at room temperature. The cell pellet is immediately resuspended in

0.7 ml sterile water and the cells are transferred to a microcentrifuge tube. Following centrifugation for 5 s at room temperature, the supernatant is discarded and the pellet frozen on dry ice. Cell pellets are stored at −80 °C until RNA isolation.

For mRNA half-life determinations using the regulatable *MET3* promoter, cells are grown in 50 ml SC-met-cys medium at 30 °C to $OD_{600} \sim 0.5$. The culture is harvested by centrifugation at $1500 \times g$ for 5 min at room temperature, and the cell pellet resuspended in 10 ml of pre-warmed (30 °C) SC-met-cys medium. Following reincubation in the shaking water bath for 5 min, P_{MET3} transcription is repressed by addition of 10 ml of pre-warmed (30 °C) SC-cys medium containing 2 mM methionine. A 2-ml aliquot is harvested immediately by a 10 s centrifugation and the cell pellet frozen on dry ice. Aliquots are then harvested at various time points after the transcription inhibition, typically 2.5, 5, 10, 20, and 40 min. Cell pellets are stored at −80 °C until RNA isolation.

For mRNA half-life determinations using strains with the temperature-sensitive *rpb1-1* allele [8, 9], encoding an altered subunit of RNA polymerase II, cells are grown in 50 ml SC medium at 25 °C to $OD_{600} \sim 0.5$. The cultures are harvested by centrifugation at $1500 \times g$ for 5 min at room temperature and the cell pellets resuspended in 10 ml of pre-warmed (25 °C) SC medium. The cultures are reincubated in a shaking water bath at 25 °C for 10 min followed by addition of 10 ml of pre-warmed (56 °C) SC medium and immediate transfer to a 37 °C shaking water bath. Aliquots are harvested as described above for the P_{MET3} system.

3.2 Isolation of Total RNA

The protocol for total RNA isolation is adapted from Herrick et al. [9], and it typically yields ~200 µg total RNA from 5 OD_{600} units of an exponentially growing yeast culture.

1. Add 500 µl of complete buffer A to the frozen cell pellet and immediately add 500 µl pre-warmed (65 °C) phenol (*see* **Note 8**).
2. Vortex for 5 s, place tube in 65 °C water bath, and incubate for 30 min. Vortex for 5 s after 30 s of incubation and then again after 5, 10, 15, 20, 25, and 30 min.
3. Place tube on ice for 5 min. Centrifuge for 2 min at maximum speed in a microcentrifuge.
4. Transfer aqueous phase to a new tube and add 400 µl phenol/chloroform/isoamyl alcohol. Vortex for 10 s and centrifuge at maximum speed for 5 min.
5. Repeat **step 4**.
6. Transfer aqueous phase to a new tube.
7. Precipitate the RNA by adding 0.1 volume 3 M NaOAc, pH 5.2, and 2.5 volumes 99.5% ethanol. Place the sample at −20 °C for at least 1 h.

8. Centrifuge at a maximum speed for 15 min and discard the supernatant.
9. Add 1 ml 80% ethanol and centrifuge at a maximum speed for 5 min. Discard supernatant, centrifuge briefly, and remove the remaining liquid using a pipette.
10. Dry the RNA pellet for ~10 min on bench and dissolve in a volume of RNase-free water that gives an RNA concentration of 1–2 μg/μl (*see* **Note 9**).
11. Determine the RNA concentration by measuring the absorbance at 260 nm (A_{260}) of an appropriate dilution (*see* **Note 10**). Freeze the RNA sample on dry ice and store at −80 °C.

3.3 Northern Blotting

The protocol for Northern blotting is adapted from He et al. [17, 18].

1. Prepare a 15 × 15 cm formaldehyde-containing 1% agarose gel as follows: Add 1.2 g agarose to 87 ml of water and boil in a microwave oven until the agarose dissolves. Cool the solution to 65 °C, add 12 ml of 10× MOPS buffer and 21 ml of formaldehyde, and mix by swirling the bottle. Cast the gel and let it solidify for 30 min.
2. Prepare the RNA samples by transferring a volume corresponding to 10 μg of total RNA to a new tube. Bring up to 10 μl with RNase-free water and add 10 μl RNA loading buffer. Incubate at 65 °C for 10 min and centrifuge briefly.
3. Load the samples on the gel, submerged in 1× MOPS electrophoresis buffer, and run at 40 V for 10 min followed by 24 V overnight. Run the gel until the bromophenol blue has migrated at least three-fourths of the gel. Place the gel on a UV transilluminator and take a picture.
4. Cut a piece of the blotting membrane to the dimensions of the gel. Pre-wet the membrane in water followed by incubation in 20× SSC for 5 min. Transfer the RNA to the membrane using the vacuum blotting system (*see* **Note 11**) essentially as described by the manufacturer (*see* **Note 12**)
5. Assess transfer efficiency by examining both the gel and membrane on a UV transilluminator. Mark the position of 25S and 18S rRNA on the membrane and cut one corner for orientation purposes.
6. Crosslink RNA to membrane using a UV crosslinker (*see* **Note 13**), set in auto-crosslink mode. Let the membrane dry on the bench. Wrap in a plastic film and store at room temperature or continue with pre-hybridization.
7. Place the membrane in a hybridization bottle with the RNA side pointing inward. Add 15 ml pre-hybridization solution and pre-hybridize with rotation in a hybridization oven for at least 3 h at 42 °C.

8. Discard the pre-hybridization solution and add 15 ml hybridization solution containing the ^{32}P-labeled probe (*see* **Note 5**). Hybridize with rotation at 42 °C for at least 16 h.
9. Transfer the membrane to a plastic box and wash with gentle agitation for 2× 10 min at room temperature and then for 2× 20 min at 58 °C, replacing the 0.1× SSC, 0.1% SDS solution between each wash step.
10. Wrap the membrane in a plastic film (*see* **Note 14**) and expose to a storage phosphor screen (*see* **Note 15**).
11. Scan and analyze the image (Figs. 1 and 2).
12. Strip the membrane by repeated incubations in boiling 1% SDS (*see* **Note 16**). Rinse the membrane briefly with 0.1× SSC 0.1% SDS and proceed to reprobing, starting with the pre-hybridization step (**step 7**).

4 Notes

1. For some genetic backgrounds, the individual knockout strains can be obtained from a deletion collection.
2. The formulation of the dropout mix varies slightly between laboratories. We are using a mix that consists of 4 g leucine, 0.5 g adenine, 0.2 g *para*-aminobenzoic acid, 1 g *myo*-inositol, and 2 g each of alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, and uracil. When making the mix for SC dropout media, the supplement of interest is left out, e.g., methionine and cysteine is left out for SC-met-cys medium.
3. The RNA loading buffer is prepared without ethidium bromide and stored as 1 ml aliquots at −20 °C. Before use, add 2.5 μl of a 10 mg/ml ethidium bromide solution and mix.
4. The pre-hybridization and hybridization solutions are prepared without salmon sperm DNA and stored at 4 °C. When preparing the hybridization solution, dextran sulfate is added as a 50% (w/v) solution. Before use, transfer 15 ml of pre-heated (42 °C) pre-hybridization/hybridization solution to a 50 ml tube. Add 150 μl of a 10 mg/ml salmon sperm DNA solution, which has previously been denatured at 95 °C for 10 min and placed on ice.
5. Radiolabeled DNA probes are prepared using the Random Primed DNA Labeling Kit (Roche) according to the manufacturer's instructions. The radiolabeled DNA is purified using mini Quick Spin DNA columns (Roche). Before the radiolabeled probe is added to the hybridization solution, it is incubated at 95 °C for 10 min and then placed on ice.

6. The volume of the medium should not exceed one-fifth of the Erlenmeyer flask volume. If starting with an exponentially growing culture, additional dilutions are typically not required.
7. Other media such as YEPD or supplemented minimal medium [16] are also frequently used. Moreover, the apparent NMD substrate status of a transcript can be influenced by the media composition [19], which means that variations of the standard formulations may be required.
8. When isolating RNA from multiple samples, the use of a repeater pipette can significantly shorten the handling time. If many samples have to be processed, the samples are divided into groups of 10–12.
9. The RNA pellet is typically dissolved in 10–15 μl water per ml of original yeast culture. Dislodge the RNA pellet from the side of the tube by carefully pipetting up and down. If the pellet is difficult to dissolve, incubate the tube at 65 °C for 5 min.
10. The A_{260}/A_{280} ratio can be used to assess the purity of the RNA sample. The pH dependence of the A_{260}/A_{280} ratio means that a dilution of a good quality RNA sample in RNase-free water will give a ratio that is lower than 2.0 (typically around 1.7).
11. An alternative approach is to transfer the RNAs by capillary blotting.
12. After partial RNA hydrolysis (50 mM NaOH, 10 mM NaCl for 5 min) and neutralization (100 mM Tris–HCl, pH 7.5 for 5 min), we allow transfer to proceed for 2 h at 40 mbar using 20× SSC as the transfer solution.
13. Make sure that the membrane is damp but not dripping wet.
14. Avoid drying the membrane as it can make stripping difficult.
15. Highly abundant transcripts such as *SCR1* typically require a short exposure (<1 h), whereas lower abundant transcripts such as *ALR1* often require at least 16 h. A Geiger counter can be used to estimate an appropriate exposure time.
16. The efficiency of stripping can be assessed by phosphorimaging.

Acknowledgments

I thank Yang Zhou, Nathan Kenney, and Tracy Nissan for the helpful comments on the manuscript. The work in my laboratory is supported by Carl Tryggers Foundation (CTS13:206), Åke Wibergs Foundation (M14-0207), and Insamlingsstiftelsen Umeå universitet.

References

1. Schweingruber C, Rufener SC, Zund D, Yamashita A, Muhlemann O (2013) Nonsense-mediated mRNA decay—mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim Biophys Acta 1829(6-7):612–623. doi:10.1016/j.bbagrm.2013.02.005
2. Kervestin S, Jacobson A (2012) NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol 13(11):700–712. doi:10.1038/nrm3454
3. He F, Jacobson A (2006) Endogenous substrates of the yeast NMD pathway. In: Maquat L (ed) Nonsense-mediated mRNA decay. Landes Bioscience, Georgetown, TX, pp 27–41
4. He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast. Mol Cell 12(6):1439–1452
5. Peccarelli M, Kebaara BW (2014) Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. Eukaryot Cell 13(9):1126–1135. doi:10.1128/EC.00090-14
6. Cherest H, Nguyen NT, Surdin-Kerjan Y (1985) Transcriptional regulation of the *MET3* gene of *Saccharomyces cerevisiae*. Gene 34(2-3):269–281
7. Mao X, Hu Y, Liang C, Lu C (2002) *MET3* promoter: a tightly regulated promoter and its application in construction of conditional lethal strain. Curr Microbiol 45(1):37–40
8. Nonet M, Scafe C, Sexton J, Young R (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Mol Cell Biol 7(5):1602–1611
9. Herrick D, Parker R, Jacobson A (1990) Identification and comparison of stable and unstable mRNAs in *Saccharomyces cerevisiae*. Mol Cell Biol 10(5):2269–2284
10. Alwine JC, Kemp DJ, Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A 74(12):5350–5354
11. Johansson MJO, Jacobson A (2010) Nonsense-mediated mRNA decay maintains translational fidelity by limiting magnesium uptake. Genes Dev 24(14):1491–1495. doi:10.1101/gad.1930710
12. Graschopf A, Stadler JA, Hoellerer MK, Eder S, Sieghardt M, Kohlwein SD, Schweyen RJ (2001) The yeast plasma membrane protein Alr1 controls Mg2+ homeostasis and is subject to Mg2+-dependent control of its synthesis and degradation. J Biol Chem 276(19):16216–16222. doi:10.1074/jbc.M101504200
13. MacDiarmid CW, Gardner RC (1998) Overexpression of the *Saccharomyces cerevisiae* magnesium transport system confers resistance to aluminum ion. J Biol Chem 273(3):1727–1732
14. Longtine MS, McKenzie A 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast 14(10):953–961
15. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294(5550):2364–2368
16. Amberg DC, Burke DJ, Strathern JN (2005) Methods in yeast genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
17. He F, Amrani N, Johansson MJO, Jacobson A (2008) Chapter 6. Qualitative and quantitative assessment of the activity of the yeast nonsense-mediated mRNA decay pathway. Methods Enzymol 449:127–147. doi:10.1016/S0076-6879(08)02406-3
18. He F, Jacobson A (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes Dev 9(4):437–454
19. Johansson MJO, He F, Spatrick P, Li C, Jacobson A (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. Proc Natl Acad Sci U S A 104(52):20872–20877. doi:10.1073/pnas.0709257105

Chapter 14

Optimizing In Vitro Pre-mRNA 3′ Cleavage Efficiency: Reconstitution from Anion-Exchange Separated HeLa Cleavage Factors and from Adherent HeLa Cell Nuclear Extract

Mihwa Na, Susana T. Valente, and Kevin Ryan

Abstract

Eukaryotic RNA processing steps during mRNA maturation present the cell with opportunities for gene expression regulation. One such step is the pre-mRNA 3′ cleavage reaction, which defines the downstream end of the 3′ untranslated region and, in nearly all mRNA, prepares the message for addition of the poly(A) tail. The in vitro reconstitution of 3′ cleavage provides an experimental means to investigate the roles of the various multi-subunit cleavage factors. Anion-exchange chromatography is the simplest procedure for separating the core mammalian cleavage factors. Here we describe a method for optimizing the in vitro reconstitution of 3′ cleavage activity from the DEAE-sepharose separated HeLa cleavage factors and show how to ensure, or avoid, dependence on creatine phosphate. Important reaction components needed for optimal processing are discussed. We also provide an optimized procedure for preparing small-scale HeLa nuclear extracts from adherent cells for use in 3′ cleavage in vitro.

Key words Cleavage and polyadenylation, Pre-mRNA processing, Pre-mRNA, 3′ end formation

1 Introduction

The site-specific cleavage of nascent pre-mRNA transcripts downstream from the stop codon is a mandatory step in the biogenesis of all eukaryotic mRNA. Cell-free systems that reconstitute pre-mRNA 3′ cleavage activity in vitro have enabled RNA biochemists to identify and study the large protein complexes responsible for this gene expression step. Early work on the pre-mRNA that receive a poly(A) tail following 3′ cleavage, that is, all pre-mRNA except those from about sixty replication dependent histone genes [1], used HeLa cell nuclear extract to reconstitute 3′ cleavage activity [2–6]. HeLa nuclear extract was subsequently fractionated to reveal a set of separable multi-subunit complexes essential for in vitro 3′ cleavage activity [7–10]. These core cleavage factors are

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_14,

cleavage polyadenylation specificity factor (CPSF, six subunits) [11, 12], cleavage stimulation factor (CstF, three subunits) [13], mammalian cleavage factors I (CF I_m, two subunits) [14, 15] and II (CF II_m, two or more subunits) [16], and, for most substrates, poly(A) polymerase (PAP) [10, 17]. Owing to the large number of essential cleavage factor polypeptides and, for several, their large size and poor solubility, the reaction has never been reconstituted using recombinant proteins. Moreover, the CF II_m fraction has not yet been fully characterized, despite earnest attempts [16, 18, 19]. The daunting complexity of the cleavage factors, and other proteins associating with them [20], poses a dilemma for biochemists seeking to learn more about the 3′ cleavage reaction: given that a recombinant system is not yet practical, and purification to homogeneity is expensive and arduous, to what extent should the cleavage factors be purified before in vitro experimentation? The answer of course depends on the nature of the experiments proposed. On the one hand, the use of unfractionated nuclear extract does not allow for the selective treatment of the cleavage factors with modifying enzymes or other reagents before reconstitution. On the other hand, over-purification risks the loss of ancillary factors that may have potentially interesting regulatory or coordinating roles via interaction with the core cleavage factors.

A compromise we have employed is to use cleavage factors fractionated from HeLa cell nuclear extract on the anion-exchange diethylaminoethyl (DEAE)-sepharose resin [21]. Several early studies began their multistep fractionation procedures with either anion-exchange [8, 10, 12, 17] or size exclusion chromatography [10, 19]. Size exclusion chromatography, using, for example, Superose 6 resin, separates poly(A) polymerase (PAP) from all other core cleavage factors, referred to collectively during early studies as the cleavage specificity factor (CSF) [10], while anion exchange separates the core factors into CPSF, CstF (co-eluting with PAP), and CF_m (denotes unseparated CF I_m and CF II_m). Thus, DEAE-sepharose anion exchange represents a simple fractionation procedure resulting in the separation of HeLa CPSF, CstF, and CF_m in vitro activities. It should be emphasized that the DEAE-separated factors are only separated from one another, not purified; they contain many co-eluting HeLa proteins, RNA [19], and, very likely, other nuclear extract constituents. Still, in combination with ammonium sulfate precipitation and a final dialysis, the DEAE-fractionated factors enable types of experiments not possible with nuclear extract (e.g., *see* ref. 21). In this report, we discuss factors that should be considered when optimizing 3′ cleavage activity using the DEAE-separated factors.

To obtain useable amounts of the partially purified cleavage factors from HeLa nuclear extract, it has been necessary to begin multistep fractionation procedures with large amounts of HeLa cells grown in suspension, typically using large flasks with spinner

agitation. For example, the first report on CF I_m, where this factor was purified to apparent homogeneity, began with 240 L of HeLa cells grown at a density of $4–6 \times 10^5$ cells/mL, which resulted in at least 312 mL of nuclear extract [14]. This volume of extract is considerable and immediately discourages experiments aiming to begin with extracts from cells transiently transfected with plasmids encoding recombinant proteins of interest. Transient transfection experiments normally reduce the scale from cells grown in liters to cells grown on 10–15 cm plates in monolayer, where the amount of extract produced is too small to begin a chromatographic separation of the factors. Nevertheless, interesting 3′ cleavage experiments can be envisioned in unfractionated nuclear extract made from transfected cells, and several methods have been reported for preparing nuclear extracts from adherent cells for in vitro 3′ cleavage activity [22–27] as well as for other nuclear activities [28–32]. Commercial kits are also available. In our experience, nuclear extracts made from plated HeLa cells tend to have lower pre-mRNA 3′ cleavage activity per extract volume than extracts made from cells grown in suspension. We have evaluated a variety of published methods for making HeLa nuclear extracts from plated cells and, in addition to our DEAE-factors discussion, describe here our current best procedure for making adherent HeLa cell nuclear extracts for use in 3′ pre-mRNA cleavage experiments.

2 Materials

It is obviously important to work using RNase-free precautions. Guidance for working RNase-free can be found in this reference [33]. The solutions of the following cleavage reaction components should be made up in diethyl pyrocarbonate (DEPC)-treated water of the highest purity available. Insofar as possible, purchase RNase-free reagents, sterilize by syringe or suction filtration, and use autoclaved plasticware.

2.1 Cleavage Reaction Component Stock Solutions (Water Based)

1. tRNA: approx. 10 mg/mL, e.g., *E. coli* MRE600.
2. 2′-dATP, pH 7–8 (in place of ATP): 100 mM (*see* **Note 1**).
3. Creatine phosphate, disodium salt (adjust for water content mass): 1 M, sterile-filtered, make fresh every 3 months or less, store in single-use aliquots at −80 °C.
4. Dithiothreitol (DTT): 200 mM (made just before use by diluting a 1 M stock (1 M stock is made in 20 mM NaOAc, pH 5.2; dilution done in water).
5. Ethylenediaminetetraacetic acid (EDTA): 25 mM, pH 8, diluted from 500 mM stock.
6. RNase inhibitor, recombinant or placental: 40 units/μL (*see* **Note 2**).

7. Polyvinyl alcohol (PVA, ≈90% hydrolyzed, avg. MW 30–70 kDa): 10% by weight in water (*see* **Note 3**).
8. Buffer D_{50}: 20% glycerol, 20 mM Hepes–Na, pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 50 mM $(NH_4)_2SO_4$, and 0.2 mM phenylmethylsulfonyl fluoride (PMSF) (*see* **Note 4**).
9. 5′-capped, radiolabeled poly(A) signal-containing RNA substrate [33]: e.g., simian virus 40 late poly(A) signal substrate, SV40L [10], or, e.g., adenovirus 2 late 3 substrate, Ad2L3 [10], or other pre-mRNA substrate, aim for about 50 nM in DEPC-treated water.

2.2 Cleavage Reaction Component Solutions (Buffer D_{50} Based)

Use DEAE-fractionated CPSF, CstF-PAP, and CF_m (*see* **Note 5**). Alternatively, HeLa cell nuclear extract from cells grown in suspension [10, 34] or from adherent cells (*see* below) may be used, 2–10 mg/mL total protein.

1. DEAE-CPSF, typically 2–7 mg/mL (total protein compared to BSA standards) dialyzed in buffer D_{50}.
2. DEAE-CstF, 2–7 mg/mL, dialyzed in buffer D_{50}.
3. DEAE-CF_m, 2–7 mg/mL, dialyzed in buffer D_{50}.
4. BSA (bovine serum albumin): 10 mg/mL (molecular biology grade, in buffer D_{50}) (*see* **Note 6**).
5. Proteinase K 2× buffer: 100 mM Tris–HCl, pH 7.5, 20 mM EDTA, 200 mM NaCl, 2% SDS.
6. Glycogen: 10 mg/mL.
7. Proteinase K: 10 mg/mL in water.
8. Saturated phenol–chloroform: equilibrated with 0.1 M Tris–HCl, pH 7.6 [35].
9. Chloroform.
10. Ammonium acetate: 10 M in water, filtered through 0.22 μm syringe filter.
11. Formamide gel loading buffer: highest purity formamide, 12 mM EDTA (diluted from 500 mM aqueous stock), 0.3% bromophenol blue, and 0.3% xylene cyanol, by weight, added dry.

2.3 Components Used in Adherent HeLa Cell Nuclear Extract Preparation

1. Dulbecco's Modification of Eagle's Medium (DMEM) with L-glutamine, 4.5 g/L glucose and sodium pyruvate.
2. Cosmic calf serum (CCS).
3. Penicillin/streptomycin, 10,000 IU/mL and 10,000 μg/mL, respectively.
4. HeLa JW36 cells (or other adherent HeLa cell type).
5. Cell culture plates: e.g., 10 cm diameter.
6. Dounce homogenizer, 7 mL, type B pestle.

7. Buffer A: 10 mM Tris–HCl, pH 8, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM PMSF.
8. Buffer C: 25% glycerol, 20 mM Tris–HCl, pH 8, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 420 mM NaCl.
9. Buffer D_{50}: 20% glycerol, 20 mM Hepes–Na, pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 50 mM $(NH_4)_2SO_4$, 0.2 mM PMSF.
10. Trypan Blue solution.
11. Dialysis membrane (MWCO 1000 Da), cut into 2 cm by 2 cm pieces.
12. Small-scale dialysis chamber: prepared by severing a capped microfuge tube near its 1.5 mL mark with a new razor blade. The space inside the lid of the tube serves as the dialysis chamber, while the ring of the remaining microfuge tube plastic serves as a clamp to hold the stretched dialysis membrane over the chamber like a drum.

3 Methods

The reconstituted in vitro cleavage reaction carried out using the DEAE-fractionated cleavage factors will be described first, followed by details on balancing the relative amounts of the cleavage factors and a discussion of three reaction components (creatine phosphate, polyvinyl alcohol (PVA) and ATP, or certain structural analogs) important for consistently efficient 3′ cleavage activity. The in vitro cleavage protocol described here is adapted from work from several laboratories [18, 36–38].

3.1 In Vitro 3′ Cleavage Reaction Using DEAE-Fractionated HeLa Cleavage Factors

The reaction is conveniently carried out in a final volume of 12.5 μL, but it may be scaled up or down as desired. Half of the volume is based on buffer D_{50} and contains either nuclear extract or the DEAE factors, all previously dialyzed in this buffer, and should be prepared entirely on ice in a 4 °C cold room using only prechilled, autoclaved plasticware. The other half is unbuffered water based and contains aqueous solutions of tRNA, ATP (or structural analog), DTT, EDTA, creatine phosphate, RNase inhibitor, polyvinyl alcohol (PVA), and the in vitro transcribed [α-^{32}P]-NTP uniformly labeled RNA substrate under study. Components that are common to all tubes in an experiment are combined in either the buffer D_{50} or in the aqueous master mix. The aqueous half of the reaction (and its master mix) should also be made on ice but can be prepared in a room at ambient temperature and then moved to the cold room. The two 6.25 μL master mix halves are mixed in a reaction tube and placed at 30 °C to start the reaction. The final buffer and salt concentrations therefore become equivalent to 0.5× buffer D_{50} (*see* **Note** 7). The volumes of buffer D_{50}

and water can be lowered in their respective mixtures to make room for other components such as small molecules, enzymes, and other variables to be tested.

This procedure describes a cleavage reaction using the DEAE-fractionated cleavage factors. If the goal is to locate cleavage factor activities following a chromatographic separation, this protocol must be adapted to that use by withholding the factor in question from the buffer D_{50} master mix tube and adding the individual column fractions to the reaction tubes in place of the withheld factor, as done, for example, in this reference [19]. In addition to being useful in locating cleavage factor activities among the fractions of a chromatographic separation, this procedure provides a control cleavage reaction starting point for experimenting with the DEAE factors. There is much room for flexibility in planning diverse experiments, individually pretreating the factors with reagents, for example, but for consistent results, it is best to aim for the same final concentrations of the water-based components listed here. For the actual cleavage factor amounts, *see* the discussion in Subheading 3.2 for balancing the cleavage factor volumes for optimal processing. Note that the volumes in the tables are the amounts *per 12.5 μL reaction* and the volumes pipetted are small. Many reaction tubes with variations in each are typically run in a single experiment; scaling up will increase the pipetting volumes to volumes that can be accurately pipetted:

1. Remove the DEAE factor aliquots to be used from −80 °C storage and place tubes on ice in the cold room.
2. Add all necessary buffer D_{50}, and the BSA, to the pre-iced buffer D_{50}-based master mix tube scaled up from the following table and leave on ice in the cold room. In scaling up, allow for one extra reaction tube (i.e., ×11 per 10 reaction tubes). Autoclaved 1.6 mL microfuge tubes are used for master mixes and reaction tubes throughout.

Buffer D_{50}-based stocks	Typical volume[a]
Buffer D_{50}	Enough to make 6.25 μL
BSA, 10 mg/mL	0.31 (250 ng)
DEAE-CPSF (4–7 mg/mL total protein)	0.4 μL
DEAE-CstF (4–7 mg/mL total protein) (*see* **Note 8**)	0.3 μL
DEAE-CF_m (4–7 mg/mL total protein)	1.2 μL
Total volume	6.25 μL (per reaction)

[a]*See* discussion below on balancing the volumes of the DEAE factor preparations

3. While the factors are slowly thawing on ice in the cold room, begin to assemble the water-based master mix tube scaled up from the table below. The indicated order of mixing is recommended but probably not critical, except that the RNase inhibitor should be added after the DTT. Mixing is done by flicking the tube gently with a finger. If following this order of addition, after adding the RNase inhibitor, but before adding the PVA, mix the tube's contents by flicking, spin briefly, and then leave the capped tube on ice.

Water-based stocks	Typical volume	Final conc.
DEPC-water	As needed	
tRNA, 10 mg/mL	0.125 µL	0.1 mg/mL
2′-dATP, 100 mM	0.25 µL	2 mM
Creatine phosphate, 1 M	0.625 µL	50 mM
DTT, 0.2 M	0.1 µL	1.6 mM
EDTA, pH 8, 25 mM	1.0 µL	2 mM
RNase inhibitor, 40 µ/µL	0.1 µL	
PVA, 10%	3.125 µL	2.5%
RNA substrate, ≈50 nM	0.5 µL	<5 nM
Total volume	6.25 µL	

4. Thaw the RNA substrate and dilute a portion to 50 nM in DEPC-treated water, heat it to 80 °C for 2 min, snap chill in an ice-water bath, vortex, spin to the bottom of the tube, and leave on ice (*see* **Note 9**).
5. Before adding the RNA to the mix, return to the cold room and mix the now-thawed DEAE factor tubes by gentle flicking and then spin them at 13,200 × *g* in a precooled microfuge located in the cold room for about 1 min.
6. Add the factors one after the other to the buffer D_{50} master mix tube, mixing by gentle flicking and then spinning briefly in the microfuge after each addition. Minimize the time outside of the ice tray. After the last factor is added, mix well and spin again in the cold microfuge for 2 min to remove any surface bubbles. Return the tube to the ice tray. Immediately refreeze any remaining cleavage factor stock by immersing the tip of the microfuge tube in liquid nitrogen (or pulverized dry ice). Mark the tube to indicate that this aliquot has been thawed and refrozen one time.
7. Continue making the aqueous master mix by adding the PVA. Before adding the PVA, vigorously mix the 10% stock

solution on a benchtop vortexer and spin the tube at 13,200 × *g* in a microfuge at room temperature for 5 min to sediment any insoluble debris (*see* **Note 10**).

8. Complete the aqueous master mix tube by adding the labeled RNA substrate. Despite the presence of the RNase inhibitor protein, the PVA allows the tube to be vortexed at medium speed without the formation of bubbles. Bring the tube to the cold room and spin at 13,200 × *g* for 1–2 min in the microfuge there.
9. On ice in the cold room, distribute 6.25 μL of the aqueous master mix into the series of reaction tubes in which the in vitro cleavage experiments will be performed.
10. Distribute 6.25 μL of the DEAE factors master mix into each tube. Mix by gentle flicking and spin for about 1 min in the cold room microfuge. (When adapting this procedure to specific experiments having liquid additions unique to the various tubes, it is possible to pipette small volumes onto the inside wall of the reaction tube and then add to the reaction mix by spinning briefly in the microfuge. In this way, the addition of different ingredients to different reaction tubes can be synchronized. This works for aqueous solutions only, where the surface tension is higher than when PVA or high protein concentrations are present.)
11. Place the tubes in a 30 °C circulating water bath for 2 h.
12. During the in vitro cleavage reaction, the denaturing polyacrylamide gel (DPAGE) [35] that will be used to separate the RNA cleavage products can be cast. In vitro cleavage does not work well on very long RNA substrates, so most are kept in the 125–250 nucleotides (nt) range. The standard SV40L and Ad2L3 substrates are in this range and a 6% DPAGE is used to resolve them.
13. Also during the in vitro cleavage incubation period, make the proteinase K 2× buffer mix. Scale up the following recipe by the number of reaction tubes, plus one.

Proteinase K mix	Volume
Water	85.5 μL
2× Proteinase K buffer	100 μL
Glycogen, 10 mg/mL	1 μL
Proteinase K, 10 mg/mL	1 μL
Total volume	187.5

14. At the end of the 2 h cleavage reaction incubation, add 187.5 µL of the proteinase K mix to each tube at room temperature, mix by inverting or flicking, spin briefly in a microfuge, and incubate at 37 °C for 15 min.
15. Add an equal volume of saturated phenol–chloroform [35], vortex strongly to completely emulsify, and then spin in a room temperature microfuge for 15 min (*see* **Note 11**).
16. Carefully transfer the top (aqueous) phase of each tube to a fresh tube containing 67 µL of 10 M ammonium acetate; mix.
17. Add 2.5 volumes (668 µL) of 95–100% ethanol, place on dry ice for 10 min or longer (*see* **Note 12**).
18. Spin 15 min in the cold room microfuge at 13,200×*g*.
19. Remove all but about 10 µL of the supernatant while carefully monitoring with a handheld radioactivity monitor (i.e., Geiger counter) to avoid discarding the pellet. Spin for 1 min and very carefully remove every last bit of visible liquid while holding the tube up to a 60 W incandescent lamp if possible while also monitoring for radioactivity (*see* **Note 13**). The glycogen usually makes the pellet visible. It may resemble a tiny piece of wet cotton. Leave the microfuge tube open on the bench for 2 min.
20. Resuspend the pellet in 8 µL formamide gel loading buffer by flicking and vortexing and then spin briefly in the microfuge to concentrate the liquid at the bottom of the tube.
21. Clamp the microfuge tubes shut with Sorenson LidLocks™ or other means, heat to 80 °C for 2 min in an aluminum microfuge tube holder block on a hotplate (or heated sand), and then snap chill on ice. Vortex strongly, spin contents to the bottom of the tube, and leave tubes at room temperature until the gel is ready to load.
22. Load equal volumes (typically 7 µL) of each sample in a preflushed well of a (typically) 6% denaturing "sequencing" electrophoresis gel, 20×20 cm size, 0.4 mm thickness (*see* **Note 14**). For the SV40L and Ad2L3 substrates cited here, run the faster xylene cyanol dye to the bottom edge, but no farther as the 3′ cleavage fragments may run off the gel. Dry the gel on Whatman gel drying paper and expose to a phosphorimager plate, or X-ray film, to image the result.

3.2 Balancing the DEAE Factor Amounts for Maximum 3′ Cleavage Activity

In practice, it can be difficult to attain consistently high efficiency when reconstituting in vitro 3′ cleavage with partially purified factors. We will focus here on important experimental details for maximizing processing in a standard 120 min in vitro cleavage reaction. We mention again the need for working under RNase-free conditions. This minimizes the background and consequently makes clearer the gel bands resulting from RNA cleavage. Another consideration, not often detailed in the literature, is the need to balance the relative amounts of

the factors for optimal activity. There is evidence that in cells some of the cleavage factors are recruited co-transcriptionally to the RNA via the RNA Pol II largest subunit's C-terminal domain (CTD) [39], and this process undoubtedly facilitates the formation of complexes with the proper stoichiometry for in vivo cleavage. In vitro, the pathway to pre-cleavage complex assembly may be different because it is not coupled to transcription. In our experience, in vitro cleavage efficiency depends on the relative amounts of the DEAE cleavage factor preparations mixed in a reaction and must be determined experimentally. In theory, quantitative Western blotting could be used to specify the stoichiometry of the factors, but this would be laborious, and the functional stoichiometry has not been determined for all factors [40, 41]. Interestingly, we have found that the ratio of the volumes is related to the still unexplained need for high concentrations of creatine phosphate to yield efficient in vitro 3′ cleavage [18].

An example of varying the DEAE factor volumes is shown in Fig. 1. The total protein concentration of these preparations was between 4 and 7 mg/mL (*see* **Note 15**). The starting point in this

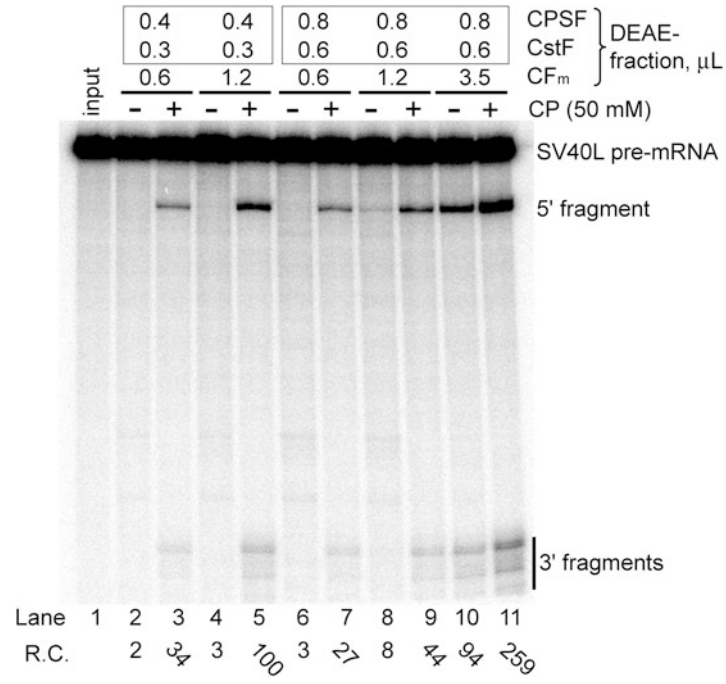

Fig. 1 In vitro 3′ cleavage of SV40L pre-mRNA: balancing the DEAE-fractionated HeLa cleavage factors for optimal activity and creatine phosphate dependence. The indicated volume in microliters (µL) of the DEAE-fractionated factors used is listed. Total protein concentrations were CPSF, 5.7 mg/mL; CstF, 6.4 mg/mL; and CF_m, 4.0 mg/mL. The upstream (5′ fragment) and downstream (3′) cleavage products are indicated. The relative cleavage (R.C., ratio of 5′ fragment to uncleaved substrate) for each reaction is normalized to lane 5 conditions. 6 % DPAGE

case was to use approximately equal volumes of CPSF and CstF and to use about twice this volume of CF_m. The relative cleavage efficiency obtained in this experiment is shown in lane 3. Holding CPSF and CstF constant and doubling CF_m led to an increase in relative cleavage (lanes 3 and 5), whereas holding CF_m constant and doubling CPSF and CstF did not significantly change the amount of cleavage products (lane 3 vs. lane 7). These results show that CF_m was the limiting factor in the first ratio chosen. At a higher volume of CPSF and CstF (0.8 and 0.6 μL, respectively), the amount of cleavage increases with the amount CF_m from 0.6 to 3.5 μL. Thus, when balancing a new batch of DEAE factors, it is CF_m that is typically limiting, and increasing its volume usually leads to more cleavage. CstF can have the opposite effect; too much can inhibit 3′ cleavage in vitro. A sound strategy to begin with is to hold CPSF (0.5 μL) and CF_m (1 μL) constant while increasing CstF from about 0.3 to 1.3 μL. This should quickly lead to the optimal ratio of CPSF to CstF (they are typically similar if their total protein content is similar). CF_m is then increased until cleavage activity levels off or no more room is available in the buffer D_{50} half of the reaction volume. However, there is one caveat to this, and it relates to creatine phosphate.

3.3 Dependence of the In Vitro Cleavage Reaction on Creatine Phosphate

The relationship to the creatine phosphate effect on in vitro cleavage activity is also illustrated in Fig. 1. In lanes 1–7, where the factors were used at comparatively low levels, detectable cleavage was clearly dependent on the addition of 50 mM creatine phosphate. But at the higher volume of CPSF (0.8 μL) and CstF (0.6 μL), increasing the CF_m from 0.6 to 3.5 μL led to some creatine phosphate-independent cleavage, as the reactions without it start to produce significant amounts of cleavage (compare lanes 6, 8, and 10). The cleavage stimulation property of creatine phosphate and related compounds may hold clues to the molecular details of 3′ cleavage [18, 42, 43]. It has at different times been postulated to be a mimic of a phosphoprotein, a mimic of the RNA Pol II CTD [42], and a serendipitous inhibitor of a 3′ cleavage-suppressing protein phosphatase [21]. How it works though is still unknown. The balancing experiment shown in Fig. 1 reveals that it can boost the activity of low CF_m concentrations just as increasing the amount of CF_m concentration can. There is evidence that the binding of CF I_m (the CF_m sub-complex containing CF I_m25 and either the 59, 68, or 72 subunit) to the RNA substrate is important for the early steps of pre-cleavage complex assembly [15]. At low CF_m concentrations, creatine phosphate may foster the proper binding of CF I_m to the RNA substrate and/or to other factors. The DEAE CF I_m fraction we used here is less pure than that from the kinetic study [15], so other explanations are certainly possible (*see* **Note 16**). Whatever the explanation, when balancing the factors for maximum in vitro cleavage activity, the nature of the

planned experiments should be taken into consideration. For example, when searching for small molecules [43] or proteins [42] that can take the place of creatine phosphate, it is obviously necessary to use a combination of factors that ensures dependence on the creatine phosphate for cleavage activity.

3.4 Other Reaction Components that Affect In Vitro 3′ Cleavage Efficiency

ATP. With the exception of the SV40L pre-mRNA substrate, ATP, or one of its structural analogs, is required for efficient in vitro cleavage [4, 18]. In fact, ATP was first thought to satisfy an energy requirement for the reaction (*see* **Note 17**), and creatine phosphate was included to allow creatine kinase to replenish the ATP pool [4]. Though the role of ATP in 3′ cleavage is still unknown, the need for its hydrolysis has been shown to be unnecessary [18]. We have found that 2′-dATP used in place of ATP leads to more efficient cleavage of the Ad2L3 substrate with the DEAE factors, and we therefore recommend its use, but most other analogs, including 3′-dATP, are less efficient than ATP. The 3′-dATP analog, cordycepin triphosphate, can nevertheless be used to replace ATP when there is a need for Mg^{2+} but when polyadenylation, which can obscure the cleavage reaction outcome, is to be avoided. The first 3′-dATP added by PAP to the cleaved 5′ fragment in the presence of Mg^{2+} cannot be extended by the polymerase, even when Mg^{2+} is present, because the 3′-OH is lacking. The 2′-dATP analog may also work in this way, though it lacks the 2′-OH, not the 3′-OH.

PVA. PVA is required for efficient in vitro cleavage and how it works is also unknown. It is sometimes assumed to be a molecular crowding agent, but we have observed that other crowding agents, like polyethylene glycol (PEG) and polyvinylpyrrolidone, cannot replace PVA. PVA is a polyol and has a dense array of alcohol groups along its chains. Interestingly, the consensus sequence of the CTD of RNA polymerase II is also very rich in alcohol side chains. Its sequence is YSPTSPS, so the consensus heptads have five hydroxyls per seven amino acid residues. The isolated recombinant CTD can stimulate 3′ cleavage in place of creatine phosphate for the Ad2L3 substrate [42, 44] (*see* **Note 18**). We have wondered if, in addition to a crowding agent role, PVA works also as a CTD mimic. However, we have no evidence for this speculation.

RNA substrate concentration. Lastly, we note that the RNA substrate should be kept below 5 nM concentration. Increasing the concentration, for example, by adding cold substrate to decrease the specific activity of the labeled substrate, decreases in vitro cleavage activity. The likely explanation is that when the RNA concentration is increased, incomplete sets of cleavage factors are distributed among different substrate molecules, statistically lowering the number of RNA substrates that can interact with all of the factors simultaneously.

3.5 In Vitro Pre-mRNA 3′ Cleavage Using HeLa Nuclear Extract from Adherent Cells

HeLa cell nuclear extract can be prepared from adherent HeLa cells for use in pre-mRNA 3′ cleavage reactions, though in our initial experience, extracts made from suspended HeLa cells generally showed greater activity. Despite the small scale used for adherent cells, which does not allow for cleavage factor separation, the use of adherent cells can be convenient because their growth can be combined with transfection, allowing for ectopic protein expression, overexpression, or RNAi knockdown experiments. Using the protocol described here, two 10 cm plates of HeLa cells grown to 80–90% confluency typically yield 60–80 μL of HeLa nuclear extract with 3′ cleavage activity comparable to that made from a large volume suspension cells (*see* Fig. 2a). In vitro 3′ cleavage carried out with HeLa nuclear extract is also to some extent dependent on creatine phosphate (*see* Fig. 2b). In our experience though, it is less dependent on creatine phosphate addition than is the reaction reconstituted from the DEAE-separated factors, and it can be difficult to find an amount of nuclear extract, from suspension cells or adherent cells, that will be completely dependent on creatine phosphate addition for detectable activity. The method we present

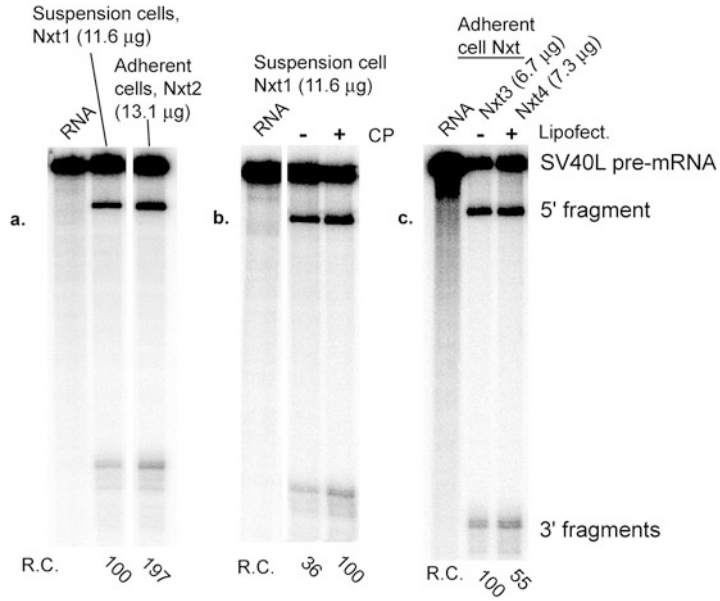

Fig. 2 In vitro 3′ cleavage activity using HeLa cell nuclear extract (Nxt) made from suspension cells and adherent cells. (**a**) Extract prepared from adherent cells as described here produces in vitro 3′ cleavage activity on par with that from extract made from suspension cells, per microgram (μg) total protein. (**b**) Creatine phosphate (CP, 50 mM) enhances cleavage activity but is not required when a typical amount (3 μL, 11.6 μg) of extract is used per 12.5 μL in vitro reaction. (**c**) HeLa cells treated with Lipofectamine 2000 (no plasmid and according to manufacturer's instructions) yielded an extract with lower cleavage activity. The SV40L pre-mRNA substrate was used in all cases. 6% DPAGE

here is in itself not new, but combines features from other published nuclear extract preparation methods to maximize in vitro 3′ cleavage activity. It should be noted that using transfection reagents may lead to decreased 3′ cleavage activity (*see* Fig. 2c), and vehicle-only control experiments should be used to gauge this effect.

Most small-scale extraction procedures are based on the classic Dignam nuclear extract preparation method [34] and involve (1) harvesting, (2) swelling, and (3) bursting the cells, followed by extraction of the nuclei at moderate salt concentration. Several protocols are available for nuclear extraction from adherent cells, mainly differing in the cell bursting technique used [25, 29–32]. To our knowledge, among these, only one method has been used to prepare in vitro 3′ cleavage activity [22–26]. The procedure we describe here is based on that method. One important modification we include is the addition of a small-scale dialysis step after salt extraction—also a typical step in the preparation of nuclear extract from suspended cells—which despite being inconvenient we found to be necessary for reliable pre-mRNA 3′ cleavage activity.

Lastly, we note that some extract-to-extract variability is unavoidable in all methods. Though we have not systematically studied it, we suspect that the precise confluency and general health of the cultured cells may be responsible for this variability. Therefore, caution should be exercised when drawing conclusions from experiments where different extracts are used within one experiment, and extensive controls and repetitions are recommended. RNAi knockdown and transient transfection are two such experiment types:

1. Grow HeLa JW36 cells (or other HeLa cell types as long as they adhere well to the plate) on two or more 10 cm plates with DMEM media supplemented with 10% CCS and 1% penicillin/streptomycin in an incubator at 37 °C and 5% CO_2. Cells should be grown until they are 80–90% confluent.
2. All steps beyond this point should be carried out in the cold room using prechilled solutions and glassware. Prepare 100 mL of PBS, 10 mL of buffer A, and 1 mL of buffer C on ice. PMSF and DTT should be added just before use.
3. Bring the cell culture plates to the cold room. Aspirate the media and wash each plate with 10 mL of PBS three times. Aspirate all the remaining PBS from the plates after leaving the plates slanted for a few seconds.
4. Add 5 mL of PBS [35] to each plate and scrape the cells off the plates using a rubber policeman.
5. Transfer the detached cells to the 15 mL falcon tube using a 10 mL pipette. Prepare a microscope slide with 2 μL of the suspension, diluting further with PBS if necessary.

6. Centrifuge at 2000 × *g* in a cold room clinical centrifuge for 5 min to collect the cells to the bottom.
7. Aspirate all the supernatant PBS and gently resuspend the cell pellet in 3 mL buffer A first by using the 10 mL pipette. Then pipette the cells up and down using a 1000 μL micropipettor to thoroughly resuspend the cells.
8. Incubate on ice for 10 min, allowing cells to swell. Verify that the cells have swollen by preparing a microscope slide with 2 μL of the suspension, and compare it to the pre-swelling sample, from **step 5**, under a light microscope. In swollen cells, the nuclei look like distinct dark spots within large translucent spheres.
9. During the swelling period, rinse the inside of the Dounce homogenizer with buffer A and place it on ice.
10. Transfer the cell suspension to the Dounce homogenizer using a 10 mL pipette. Carefully insert the B-type pestle into the Dounce, avoiding bubbles, and slowly move it up and down 20 times, on ice (*see* **Note 19**).
11. Check for cell lysis by taking 2 μL of the cells in buffer A and mixing in a microfuge tube with an equal volume of Trypan Blue. Observe cell lysis under the microscope. The lysed cells appear blue and ~90% of the cells should be lysed. If they are not, a few more Dounce strokes can be added until most of the cells are lysed. Transfer the lysate back to the 15 mL falcon tube.
12. Centrifuge at 4000 × *g* for 10 min and aspirate the supernatant away from the nuclei.
13. Estimate the packed nuclear volume (PNV) either by using the graduation on the falcon tube or using a separate empty falcon tube and water.
14. Slowly resuspend the nuclear pellet in 1.5 PNV of buffer C, using a 200 μL micropipettor. This results in a ~250 mM NaCl concentration for the nuclear extraction.
15. Transfer the nuclear mixture to a microfuge tube. Incubate for 30 min on a Nutator mixer.
16. During this time, wash the dialysis membrane in DEPC-treated water, and leave it submerged in a 50 mL falcon tube with DEPC-treated water.
17. Centrifuge the nuclei suspension at 13,200 × *g* for 15 min.
18. Transfer the supernatant (the nuclear extract) to the small-scale dialysis chamber. Lay the dialysis membrane over the top and then seal the "drum" by sliding the ring of cut tube over the opening. Invert the chamber and make sure the liquid moves into contact with the membrane. Place the small-scale dialysis chamber in a beaker containing 250 mL of buffer D_{50}, the membrane facing down and in contact with the buffer (no trapped air).

19. Dialyze with gentle stirring for 2 h.
20. Remove the dialysis chamber from buffer D_{50} and place on the bench, the membrane facing up. Blot away any excess buffer sitting on the membrane. Using a new razor, cut a slit into the dialysis membrane. Using a P200 micropipettor, collect as much of the dialyzed nuclear extract as possible and deposit into a prechilled microfuge tube (*see* **Note 20**).
21. Use 2 μL to measure the total protein concentration via Bradford assay, compared to standard solutions of a protein such as BSA. Starting from two 10 cm plates of HeLa cells, this procedure typically yields 60–80 μL of nuclear extract having a protein concentration of 1–2 mg/mL.
22. Aliquot the rest of the extract in microfuge tubes. Snap freeze the aliquots in liquid nitrogen and store at −80 °C. Avoid repeated freeze-thaw cycles.
23. For in vitro cleavage reactions, use 3–6.25 μL of the nuclear extract in place of the DEAE cleavage factors as described in Subheading 3.1.

4 Notes

1. We have surveyed a variety of ATP structural analogs and find that 2′-dATP is the most effective [46].
2. RNase inhibitors work against a type of ribonuclease unrelated to the 3′ cleavage endoribonuclease; there is no danger of inhibiting 3′ cleavage.
3. Dissolve PVA with rocking on Nutator at room temperature over 1–2 days, vortex, and spin in microfuge at 13,200 × *g* for 5 min just before each use.
4. Add PMSF, a serine protease inhibitor, just before contact with factors when possible. It has a short half-life in water.
5. DEAE-sepharose and Mono Q are the two anion-exchange resins that have been used the most to separate the cleavage factor activities. For DEAE-sepharose chromatographic procedures, see the following references [8, 10, 12, 17, 38, 44].
6. BSA prevents the total protein concentration from falling too low. 250–500 ng is typically used per 12.5 μL reaction. DEAE factors work well with 250 ng. To be safe, more highly purified factors, which have less total protein, should be supplemented by higher amounts of BSA.
7. This results in a buffer concentration of only 10 mM, while the creatine phosphate concentration is 50 mM. Fortunately, the creatine phosphate pH is close to that of the buffer and does

not change the pH. If acidic or basic components are used in adaptations of this protocol, take care to keep the pH in the 7.5–8.0 range.

8. CstF and PAP elute in the void volume of anion-exchange columns such as DEAE and Mono Q. It is therefore not necessary to supplement the DEAE factors with recombinant PAP when processing PAP-dependent substrates, i.e., all substrates except the SV40L pre-mRNA. If PAP is suspected to be low, 2 ng recombinant bovine PAP can be used [45].
9. Though heat denaturation is probably not critical for most substrates, it is done to break up any unexpectedly strong intramolecular secondary structure that might be present and affect processing. Longer substrates have more chance for secondary structure and may work poorly in vitro.
10. 10% PVA is normally stored at −20 °C and can be thawed quickly in a warm water bath. PVA is viscous and difficult to pipette accurately. Draw it slowly into the pipette tip to avoid air bubbles, and gently pump after dispensing into the mixture to rinse the inside of the pipette tip. The benchtop microfuge used throughout this method was an Eppendorf 5415D. Its top speed is 13,200 × *g*.
11. Excessive insoluble material at the phase interface may indicate (1) the PVA solution is too old or too concentrated, (2) tubes are made out of material other than polypropylene and react with chloroform, and (3) a need for longer microfuge centrifugation time to separate phases.
12. It is safe to stop the procedure at this step and leave the precipitation tubes at −20 °C or lower overnight.
13. This procedure can be tedious for a large number of samples but it is critical since the pellet can be lost or fracture and be partially lost. For best results, remove as much supernatant as possible with a P1000 micropipettor, spin again, remove all but 10 μL, spin again, and remove every last trace. If this is done, there is no need to rinse the pellet with 70% aqueous ethanol. If the salt in the supernatant is not removed at this step, migration through the gel may be adversely affected.
14. One of the plates can be siliconized [35] to help the gel adhere to only one of the plates when the gel is removed from the glass plates. Since sequencing gels are now rarely used, the spacers and gel combs to form the square wells can be hard to find. Labrepco is currently selling them under SKU: 21035043.
15. Bio-Rad Bradford reagent method was used, with comparison to BSA standards.
16. RNA Pol II can replace creatine phosphate in in vitro cleavage [42], and we have detected some RNA Pol II in CF_m by

Western blotting. Though it seems not to be in high enough concentration to be responsible for this creatine phosphate-independent cleavage, we cannot rule it out.

17. The hydrolysis of esters such as RNA and DNA in water is thermodynamically favorable. Thus, no ATP should be required. However, energy could be required to assemble the proper cleavage factor complexes needed to carry out cleavage at a specific site.
18. Along with the partially purified core cleavage factors, the recombinant RNA Pol II CTD, in the form of a GST-CTD fusion protein, can stimulate in vitro cleavage in place of creatine phosphate for the Ad2L3 substrate but not for the SV40L substrate [42, 44]. In our experience, it cannot replace creatine phosphate in unfractionated HeLa cell nuclear extract with either substrate.
19. As an alternative to douncing, cells can also be passed through syringe needle [30]. However, using a syringe needle tends to generate bubbles. The detergents such as NP-40 and digitonin have also been used [31, 32]. The extent to which cells are exposed to the detergent must be empirically determined. Repeated freeze-thaw cycles (three times) can also lyse the cells, but this method can be time consuming [28].
20. The plastic ring may be removed and the chamber, with its loose membrane still over the top, may be inserted into a new microfuge tube whose cap has been removed. Brief spinning in a microfuge will transfer any remaining extract through the slit and into the tube. There may also be commercially available small-scale dialysis alternatives.

Acknowledgments

This work was supported by grant 5SC1GM083754 to K.R. from the National Institutes of Health. Additional infrastructural support at the City College of New York was provided by the NIH National Center for Research Resources (2G12RR03060-26A1) and the National Institute on Minority Health and Health Disparities (8G12MD007603-27).

References

1. Albig W, Doenecke D (1997) The human histone gene cluster at the D6S105 locus. Hum Genet 101(3):284–294
2. Conway L, Wickens M (1987) Analysis of mRNA 3′ end formation by modification interference: the only modifications which prevent processing lie in AAUAAA and the poly(A) site. EMBO J 6(13):4177–4184
3. Gilmartin GM, McDevitt MA, Nevins JR (1988) Multiple factors are required for specific RNA cleavage at a poly(A) addition site. Genes Dev 2(5):578–587

4. Moore CL, Sharp PA (1985) Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell 41(3):845–855
5. Sperry AO, Berget SM (1986) In vitro cleavage of the simian virus 40 early polyadenylation site adjacent to a required downstream TG sequence. Mol Cell Biol 6(12):4734–4741
6. Zhang F, Cole CN (1987) Identification of a complex associated with processing and polyadenylation in vitro of herpes simplex virus type 1 thymidine kinase precursor RNA. Mol Cell Biol 7(9):3277–3286
7. Christofori G, Keller W (1988) 3′ cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. Cell 54(6):875–889
8. Gilmartin GM, Nevins JR (1989) An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev 3(12B):2180–2190
9. McLauchlan J, Moore CL, Simpson S, Clements JB (1988) Components required for in vitro cleavage and polyadenylation of eukaryotic mRNA. Nucleic Acids Res 16(12):5323–5344
10. Takagaki Y, Ryner LC, Manley JL (1988) Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell 52(5):731–742
11. Bienroth S, Wahle E, Suter-Crazzolara C, Keller W (1991) Purification of the cleavage and polyadenylation factor involved in the 3′-processing of messenger RNA precursors. J Biol Chem 266(29):19768–19776
12. Murthy KG, Manley JL (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J Biol Chem 267(21):14804–14811
13. Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev 4(12A):2112–2120
14. Ruegsegger U, Beyer K, Keller W (1996) Purification and characterization of human cleavage factor Im involved in the 3′ end processing of messenger RNA precursors. J Biol Chem 271(11):6107–6113
15. Ruegsegger U, Blank D, Keller W (1998) Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol Cell 1(2):243–253
16. de Vries H, Ruegsegger U, Hubner W, Friedlein A, Langen H, Keller W (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J 19(21):5895–5904
17. Christofori G, Keller W (1989) Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. Mol Cell Biol 9(1):193–203
18. Hirose Y, Manley JL (1997) Creatine phosphate, not ATP, is required for 3′ end cleavage of mammalian pre-mRNA in vitro. J Biol Chem 272(47):29636–29642
19. Takagaki Y, Ryner LC, Manley JL (1989) Four factors are required for 3′-end cleavage of pre-mRNAs. Genes Dev 3(11):1711–1724
20. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL (2009) Molecular architecture of the human pre-mRNA 3′ processing complex. Mol Cell 33(3):365–376. doi:10.1016/j.molcel.2008.12.028
21. Ryan K (2007) Pre-mRNA 3′ cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol 4(1):26–33
22. Cevher MA, Zhang X, Fernandez S, Kim S, Baquero J, Nilsson P, Lee S, Virtanen A, Kleiman FE (2010) Nuclear deadenylation/polyadenylation factors regulate 3′ processing in response to DNA damage. EMBO J 29(10):1674–1687. doi:10.1038/emboj.2010.59
23. Di Giammartino DC, Li W, Ogami K, Yashinskie JJ, Hoque M, Tian B, Manley JL (2014) RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3′ UTRs. Genes Dev 28(20):2248–2260. doi:10.1101/gad.245787.114
24. Di Giammartino DC, Shi Y, Manley JL (2013) PARP1 represses PAP and inhibits polyadenylation during heat shock. Mol Cell 49(1):7–17. doi:10.1016/j.molcel.2012.11.005
25. Kleiman FE, Manley JL (2001) The BARD1-CstF-50 interaction links mRNA 3′ end formation to DNA damage and tumor suppression. Cell 104(5):743–753
26. Kleiman FE, Wu-Baer F, Fonseca D, Kaneko S, Baer R, Manley JL (2005) BRCA1/BARD1 inhibition of mRNA 3′ processing involves targeted degradation of RNA polymerase II. Genes Dev 19(10):1227–1237. doi:10.1101/gad.1309505
27. Nazeer FI, Devany E, Mohammed S, Fonseca D, Akukwe B, Taveras C, Kleiman FE (2011) p53 inhibits mRNA 3[prime] processing through its interaction with the CstF/BARD1 complex. Oncogene 30(27):3073–3083
28. Coqueret O, Gascan H (2000) Functional interaction of STAT3 transcription factor with the cell cycle inhibitor p21WAF1/CIP1/SDI1. J Biol

Chem 275(25):18794–18800. doi:10.1074/jbc.M001601200

29. Folco EG, Lei H, Hsu JL, Reed R (2012) Small-scale nuclear extracts for functional assays of gene-expression machineries. J Vis Exp (64). doi:10.3791/4140
30. Lee KA, Bindereif A, Green MR (1988) A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. Gene Anal Tech 5(2):22–31
31. Liu X, Fagotto F (2011) A method to separate nuclear, cytosolic, and membrane-associated signaling molecules in cultured cells. Sci Signal 4(203):pl2. doi:10.1126/scisignal.2002373
32. Schreiber E, Matthias P, Muller MM, Schaffner W (1989) Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. Nucleic Acids Res 17(15):6419
33. Chabot B (1994) Synthesis and purification of RNA substrates. In: Higgins SJ, Hames BD (eds) RNA processing, a practical approach, vol 1. Oxford University Press, Oxford, pp 1–30
34. Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res 11(5):1475–1489
35. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
36. Gilmartin GM (1997) In vitro analysis of mammalian cell mRNA 3′ processing. In: Richter JD (ed) MRNA formation and function. Academic, New York, pp 79–98
37. Moore CL (1990) Preparation of mammalian extracts active in polyadenylation. Methods Enzymol 181:49–74
38. Wahle E, Keller W (1994) 3′ end-processing of mRNA. In: Higgins SJ, Hames BD (eds) RNA processing: a practical approach, vol 2. Oxford University Press, Oxford UK, pp 1–33
39. Bentley DL (2014) Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15(3):163–175. doi:10.1038/nrg3662
40. Bai Y, Auperin TC, Chou CY, Chang GG, Manley JL, Tong L (2007) Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol Cell 25(6):863–875. doi:10.1016/j.molcel.2007.01.034
41. Yang Q, Coseno M, Gilmartin GM, Doublie S (2011) Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure 19(3):368–377. doi:10.1016/j.str.2010.12.021
42. Hirose Y, Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. Nature 395(6697):93–96
43. Ryan K, Khleborodova A, Pan J, Ryan XP (2009) Small molecule activators of pre-mRNA 3′ cleavage. RNA 15(3):483–492. doi:10.1261/rna.1262509
44. Ryan K, Murthy KG, Kaneko S, Manley JL (2002) Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3′ cleavage. Mol Cell Biol 22(6):1684–1692
45. Wahle E (1991) Purification and characterization of a mammalian polyadenylate polymerase involved in the 3′ end processing of messenger RNA precursors. J Biol Chem 266(5):3131–3139
46. Khleborodova A, Pan X, Nagre NN, Ryan K. 2016. An investigation into the role of ATP in the mammalian pre-mRNA 3' cleavage reaction. Biochimie 125:213–222.

Chapter 15

Unbiased Interrogation of 3D Genome Topology Using Chromosome Conformation Capture Coupled to High-Throughput Sequencing (4C-Seq)

Rutger W.W. Brouwer, Mirjam C.G.N. van den Hout, Wilfred F.J. van IJcken, Eric Soler, and Ralph Stadhouders

Abstract

The development and widespread implementation of chromosome conformation capture (3C) technology has allowed unprecedented new insight into how chromosomes are folded in three-dimensional (3D) space. 3C and its derivatives have contributed tremendously to the now widely accepted view that genome topology plays an important role in many major cellular processes, at a chromosome-wide scale, but certainly also at the level of individual genetic loci. A particularly popular application of 3C technology is to study transcriptional regulation, allowing researchers to draw maps of gene regulatory connections beyond the linear genome through addition of the third dimension. In this chapter, we provide a highly detailed protocol describing 3C coupled to high-throughput sequencing (referred to as 3C-Seq or more commonly 4C-Seq), allowing the unbiased interrogation of genome-wide chromatin interactions with specific genomic regions of interest. Interactions between spatially clustered DNA fragments are revealed by cross-linking the cells with formaldehyde, digesting the genome with a restriction endonuclease and performing a proximity ligation step to link interacting genomic fragments. Next, interactions with a selected DNA fragment are extracted from the 3C library through a second round of digestion and ligation followed by an inverse PCR. The generated products are immediately compatible with high-throughput sequencing, and amplicons from different PCR reactions can easily be multiplexed to dramatically increase throughput. Finally, we provide suggestions for data analysis and visualization.

Key words Chromosome conformation capture (3C), 4C-Seq, 4C-Seq, Genome-wide, Long-range gene regulation, Chromatin looping, DNA, Bioinformatics

1 Introduction

Chromosome conformation capture (3C) and high-throughput derivatives (4C, 3C-/4C-Seq, Capture-C, T2C, Hi-C, Capture-Hi-C, ChIA-PET) enable the reconstruction of average topological genome conformations from populations of cells [1, 2]. This knowledge is essential to fully understand gene regulatory principles, as it has been clearly demonstrated that gene transcription is

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_15,

intimately linked to three-dimensional (3D) genome organization. Indeed, gene regulatory elements may interact with their target genes over large genomic distances (hundreds to thousands of kilobases (kb)) via chromatin looping mechanisms (reviewed in [3]) to modulate the transcriptional activity of target promoters. Therefore, defining long-range chromatin interactions and their dynamics across different experimental settings has the potential to yield critical insights into gene regulatory processes. 3C, for example, allows researchers to detect gene distal cis-regulatory elements through their physical co-association with target promoters in large and complex genomes. 3C also represents an essential tool for studying genotype–phenotype relationships in mammals and the functional impact of common noncoding genomic variants in humans (i.e., [4–9]).

The principle underlying 3C is one of elegant simplicity. 3C (and all 3C-related technologies) relies on proximity ligation-mediated capture of chromatin interactions (Fig. 1). Typically, formaldehyde crosslinking is used to fix the native 3D genome organization, followed by restriction enzyme digestion of the genome. This way, distal chromatin fragments that were in close proximity in the nuclear space at the time of fixation remain physically linked and can be ligated to each other. Subsequent PCR strategies and deep sequencing approaches are used to detect these ligation products and to reconstruct the 3D chromatin conformation of the cells under investigation.

We and others [10] have used 3C-sequencing (called 4C-Seq or more commonly 4C-Seq), a high-throughput 3C derivative to analyze topological conformations of genomic regions of interest with genome-wide coverage. 4C-Seq uses a single chosen genomic region of interest (e.g., enhancer, promoter, domain boundary; referred to as the "viewpoint") and interrogates the whole genome for chromatin co-associations with this region. As such, 4C-Seq by itself is not truly a genome-wide assay as it is focused on chosen genomic regions. To overcome this limitation, we apply 4C-Seq in a multiplexed fashion to enable analysis of dozens of viewpoints in parallel with genome-wide coverage [11].

2 Materials

All materials, reagents, and software listed are divided using the same subheadings as used in Subheading 3 describing the methods. All buffers and solutions described below are prepared using nuclease-free deionized water. Reagents can be stored at room temperature unless stated otherwise.

2.1 Design of a 4C-Seq Experiment

1. Computer with standard restriction analysis/PCR primer design software.

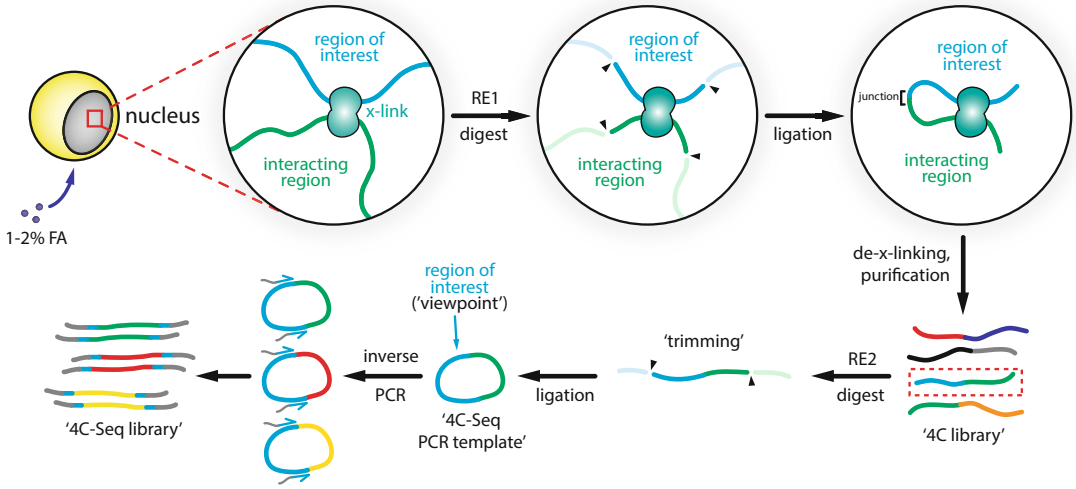

Fig. 1 Principles underlying the 4C-Seq protocol. Depicted is a schematic of the different experimental steps required to obtain a 4C-Seq library. Cells are treated with formaldehyde ("FA"; 1–2 %), causing nuclear architecture to be stably fixed through the formation of protein–protein and protein–DNA crosslinks (shown as an "x-link"). As a result, interacting genomic regions normally separated on the linear chromosome (exemplified by the *blue* and *green* strands) are maintained in close proximity during chromatin fragmentation (digestion with the primary restriction enzyme, "RE1") and proximity ligation. The latter step creates chimeric molecules (the "3C library") with a ligation junction that allows their identification. In 3C-/4C-Seq protocols, 3C libraries are digested a second time (with RE2) to trim the average size of the individual molecules, allowing for efficient PCR amplification after self-circularization through a second ligation. PCR primers are designed on a region of interest (the "viewpoint," depicted in *blue* here) and face outward to amplify all unknown fragments that have been ligated to the viewpoint (represented by the *green*, *red*, and *yellow* fragments). The addition of sequencing adapters to the PCR primers (shown as *gray* overhangs) allows for direct high-throughput sequencing of 4C-Seq libraries

2.2 Crosslinking of Cells and Isolation of Nuclei

1. Mammalian cells of interest (i.e., cultured cell lines or primary cells, FACS purified cells, or freshly isolated tissue samples) (*see* **Note 1**).
2. Cell strainer (40 μM or 70 μM depending on the cell size).
3. Appropriate cell culture medium or 10 % heat-inactivated FCS in PBS (FCS/PBS).
4. 37 % formaldehyde (*Toxic:* handle in chemical fume hood) (*see* **Note 2**).
5. 1-M glycine in water (0.22 μM filtered).
6. Protease inhibitors (e.g., complete protease inhibitors, Roche).
7. Freshly prepared cell lysis buffer (10-mM Tris–HCl, pH 8.0, 10-mM NaCl, 0.2 % NP-40 containing protease inhibitors), store on ice until needed (*see* **Note 3**).

2.3 Generation of the 3C Library: Digestion and Ligation

1. Primary restriction enzyme and appropriate buffer (*see* **Note 4**).
2. 10 % (wt/vol) SDS.
3. 20 % (wt/vol) Triton X-100.

4. Proteinase K (10 mg/ml).
5. Basic agarose gel electrophoresis equipment and reagents.
6. 10× T4 DNA ligation buffer (Roche) (*see* **Note 5**).
7. T4 DNA ligase (highly concentrated, 5 U/μl, Roche).

2.4 De-crosslinking and DNA Purification

1. Phenol/chloroform/isoamyl alcohol (25:24:1), saturated with 100-mM Tris–HCl, pH 8.0 (*Toxic:* handle in chemical fume hood).
2. DNase-free RNase A (10 mg/ml).
3. 2-M sodium acetate, pH 5.6 or 3-M sodium acetate, pH 5.2.
4. 100% and 70% ethanol.
5. 10-mM Tris–HCl, pH 7.5.

2.5 Generation of the 4C-Seq PCR Template: Digestion and Ligation

1. Secondary restriction enzyme and appropriate buffer (*see* **Note 6**).
2. Glycogen (20 mg/μl), molecular biology grade.

2.6 DNA Purification, PCR Amplification, and 4C-Seq Library Preparation

1. Spin column-based DNA purification kit (*see* **Note 7**).
2. DNA quantification system (e.g., spectrophotometric or fluorometric).
3. Expand Long Template PCR System (Roche).
4. dNTPs (10 mM).
5. Viewpoint-specific inverse PCR oligonucleotides.
6. High Pure PCR Product Purification kit (Roche) or AMPure XP beads (Beckman Coulter) (*see* **Note 8**).

2.7 High-Throughput Sequencing and Data Analysis

1. Bioanalyzer and DNA 12000 chip cartridges (Agilent) (*see* **Note 9**).
2. Accurate DNA quantification methodology.
3. Illumina high-throughput sequencing system and required reagents.

2.8 Primary Data Analysis and Downstream Data Analysis (*See* Note 26)

1. *SAMtools* available at http://www.htslib.org/ [12].
2. *BEDTools* available at https://github.com/arq5x/bedtools2 [13].
3. *Bowtie* available at http://bowtie-bio.sourceforge.net/ [14].
4. Illumina base-calling software (*bcl2fastq*) available from http://www.illumina.com/.
5. *IGV genome browser* available at http://www.broadinstitute.org/igv/ [15].
6. *Python* available at http://www.python.org/. All simple 4C-Seq analysis scripts have been developed using Python release 2.x.

7. *Pysam* available at https://github.com/pysam-developers/pysam.
8. 4C-Seq analysis scripts available at https://github.com/RWWB/simple3Cseq.
9. FastA file with the reference genome sequence.
10. Chromosome sizes for the reference genome (generated from FastA).
11. Bowtie index for the reference genome (generated from FastA).
12. Software packages for in-depth data analysis (e.g., *r3Cseq*, *Basic4CSeq*, *fourSig*, *4Cseqpipe*, *FourCSeq*).

3 Methods

3.1 Design of a 4C-Seq Experiment

1. Determine "bait(s)" or "viewpoint(s)" of interest to be analyzed. Typically, these represent discrete genomic regions defined by a known or suspected regulatory function (e.g., enhancer or promoter)—however, any genomic region is in principle amenable to 4C-Seq analysis.
2. Select a (set of) suitable primary restriction enzyme(s) based on (1) general performance in 3C-type conditions, (2) insensitivity to mammalian CpG methylation, (3) no/very low star activity, (4) excellent (>95%) ligation efficiencies after digestion, and (5) recognition site distribution surrounding the desired viewpoint(s) and the desired resolution (*see* **Note 4** and Fig. 2).
3. Select suitable secondary restriction enzymes based on (1) insensitivity to mammalian CpG methylation, (2) excellent (>95%) ligation efficiencies after digestion, and (3) compatibility with the primary restriction enzyme in generating a suitable fragment for inverse PCR primer design (*see* Fig. 2) (*see* **Note 6**).
4. Design inverse PCR primer pairs on restriction fragment ends of the viewpoint(s) of interest (*see* Fig. 2) (*see* **Note 10**).

3.2 Crosslinking of Cells and Isolation of Nuclei

1. Collect cells or tissues; resuspend in fresh culture medium or FCS/PBS. If required (e.g., when using tissue samples), prepare a single-cell suspension by filtering through a cell strainer (*see* **Note 1**).
2. Determine cell concentration and further dilute the cell suspension as necessary using fresh culture medium or FCS/PBS (*see* **Note 11**). We prefer to crosslink cells at a 1×10^6 cells per ml density to ensure standardized conditions (e.g., 10×10^6 cells in 10 ml of the medium). We have conducted successful 4C-Seq experiments using this protocol starting from $10–20 \times 10^6$ to as little as $1–0.5 \times 10^6$ cells [16–18].

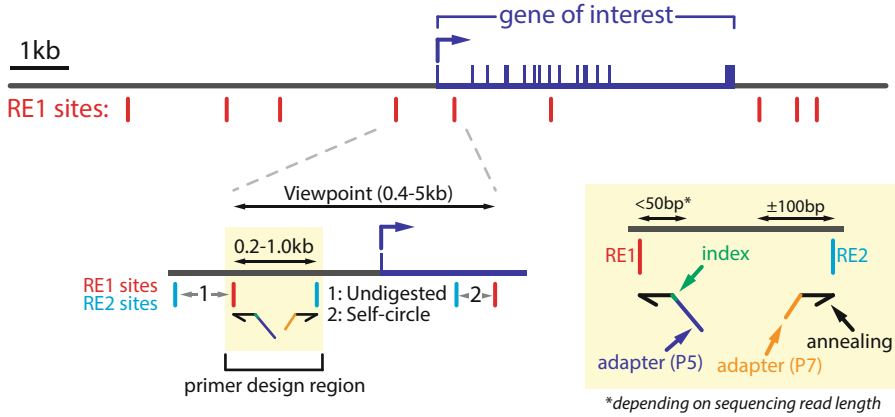

Fig. 2 Primer design considerations. A hypothetical locus of interest is shown to illustrate several aspects of 4C-Seq primer design. Primary restriction enzyme (RE1) sites are scattered around the locus, determining local resolution of the 4C-Seq experiment. Viewpoint fragment length is considered optimal between 0.4 and 5.0 kb and is generally chosen to be located as close as possible to (or preferably overlapping with) the specific genomic element of interest—in this case the promoter region of a chosen gene (the transcription start site is depicted as an *arrow*, exons as *vertical blue lines*). Selected viewpoint fragments are then examined for secondary restriction enzyme (RE2) sites. Primers can be designed at both the 5′- and 3′-ends of the viewpoint fragment (as delineated by the RE1 sites), as long as the final RE1–RE2 fragment is >200 bp (thus ensuring efficient self-circularization) and allows primer design near the RE1 and RE2. In the example, the 5′ RE1–RE2 combination was chosen for primer design (*see* magnification marked in *yellow*). Sequencing is initiated from the RE1 side of the viewpoint; hence, the corresponding primer should contain an optional index and the P5 Illumina sequencing adapter as overhang. This "reading primer" should be positioned as close to the RE1 site as possible (preferably encompassing it) to ensure enough sequence read length is left for aligning the unknown interacting sequence to the genome. The design of the nonreading primer with P7 adapter overhang near the RE2 site allows for more flexibility, as sequencing is not initiated from this side. Inherent to the experimental strategy of 3C-/4C-Seq, two fragments will be amplified with a much higher frequency (also *see* Fig. 3): (1) the "undigested" fragment due to a failure to cut the RE1 site next to the reading primer (the RE1–RE2 fragment directly adjacent to the viewpoint RE1–RE2 fragment) and (2) the "self-circle" fragment due to the self-circularization of the viewpoint fragment in the first round of ligation (the RE1–RE2 fragment at the opposite end of the viewpoint fragment not used for primer design)

3. Add 37% formaldehyde directly to the cell suspensions to a final concentration of 1–2% (*see* **Note 12**). Immediately mix by inverting and incubate at room temperature for 10 min under rotation.
4. Transfer the cells to ice and add glycine to a final concentration of 0.125 M.
5. Immediately centrifuge the crosslinked cells for 8 min at 340 × *g* (4 °C) and remove all supernatant.
6. Gently resuspend the cell pellet in 5 ml of prechilled cell lysis buffer. Incubate for 10–15 min on ice.
7. Centrifuge the mixture for 5 min at 650 × *g* (4 °C) to pellet the nuclei.

8. Pelleted nuclei can be washed with PBS and transferred to 1.5-ml microcentrifuge tubes (e.g., as aliquots), snap-frozen in liquid N_2, and stored at −80 °C for at least a year.

3.3 Generation of the 3C Library: Digestion and Ligation

1. Gently resuspend the nuclei in 0.5 ml of 1.2× restriction buffer. Do not use more than 10×10^6 nuclei per individual reaction (*see* **Note 13**).
2. Place the tubes in a thermomixer at 37 °C and add 15 μl of 10% SDS (final, 0.3%).
3. Incubate at 37 °C for 1 h while shaking at 900 rpm.
4. Add 50 μl of 20% Triton X-100 (final, 2%) (*see* **Note 14**).
5. Incubate at 37 °C for 1 h while shaking at 900 rpm.
6. Remove a small (5–10 μl) aliquot from each sample and store it overnight at −20 °C. The DNA extracted from these nuclei will be used to determine the digestion efficiency (*see* **steps 9** and **10**).
7. Add 400 U of the preferred restriction enzyme to the remaining sample and incubate overnight at 37 °C while shaking at 900 rpm (*see* **Note 15**).
8. Remove a small (5–10 μl) aliquot from each sample. The DNA extracted from these nuclei will be used to determine the digestion efficiency (*see* **steps 9** and **10**).
9. Add 10 μl of proteinase K to the two control aliquots and add 10-mM Tris–HCl, pH 7.5 to a final volume of 100 μl. Incubate at 65 °C for 1 h to reverse formaldehyde crosslinks.
10. Directly run a 20 μl aliquot of the control samples alongside each other on a 0.6% (wt/vol) standard agarose gel. The DNA from undigested control samples should run as a discrete high molecular weight band (>12 kb), while a DNA smear should appear after a successful digestion (Fig. 3) (*see* **Note 16**).
11. If digestion was successful, heat-inactivate the restriction enzyme by incubating at 65 °C for 20 min. If the enzyme cannot be heat-inactivated, add 80 μl of 10% SDS (final, 1.6%) before incubating at 65 °C for 20 min.
12. Transfer the sample to a 50-ml centrifugation tube and add 6.125 ml of 1.15× ligation buffer (*see* **Note 17**).
13. Only if SDS was added in **step 11**, add 375 μl of 20% Triton X-100 (final, 1%) and incubate for 1 h at 37 °C in a water bath.
14. Add 100 U of T4 DNA ligase and incubate at 16 °C for 4 h or overnight.

3.4 De-crosslinking and DNA Purification

1. Add 30 μl of proteinase K and incubate for at least 4 h (overnight is also possible) at 65 °C to de-crosslink the samples.
2. Add 30 μl of RNase A and incubate for 30–45 min at 37 °C.

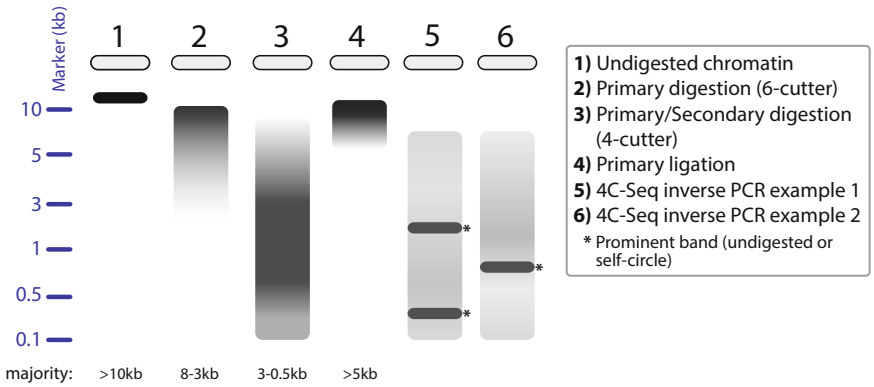

Fig. 3 4C-Seq wet-lab quality control. Artificial representation of typical DNA smears obtained after agarose gel electrophoresis during the various quality control steps of the 4C-Seq protocol (typical average sizes of the majority of DNA fragments are shown below the "gel"). Undigested chromatin (1) runs as a sharp high molecular weight band. After the first digestion (2 + 3), this sharp band should largely dissolve into a downward smear (with an average size depending on whether a 6-bp- or 4-bp-recognizing enzyme was selected). After the first ligation (4), the DNA fragment size returns upward as a compact high molecular weight smear/band. The second round of digest (3)—using a frequent cutting 4-bp-recognizing enzyme—results again in a smear of low molecular weight fragments. The second ligation (the self-circularization step) does not result in visible DNA fragment size changes on the gel. Inverse PCR (5 + 6) typically produces a wide range of DNA fragments (representing detectable interacting fragments) that appear as a smear on the gel. Two prominent bands (the "undigested" and "self-circle" fragments, *see* Fig. 2 for an explanation of their origin) are often readily visible (5). Note that these fragments can be very small (<100 bp) and therefore they do not always appear on the gel (6)

3. Cool the samples to room temperature and add 7 ml of phenol/chloroform/isoamyl alcohol and shake the samples vigorously.
4. Centrifuge for 15 min at 3200 × *g*.
5. Transfer the upper aqueous phase into a new 50-ml tube. Add 7 ml of Milli-Q H_2O and 1.5 ml of sodium acetate and add 35 ml of 100 % ethanol.
6. Mix thoroughly and place the sample at −80 °C for 2–3 h until the liquid is frozen solid.
7. Directly centrifuge the frozen samples for 45 min at 3200 × *g* (4 °C).
8. Remove the supernatant and add 10 ml of 70 % ethanol.
9. Centrifuge the mixture for 15 min at 3200 × *g* (4 °C).
10. Remove the supernatant, air-dry, and dissolve the pellet in 150 µl of 10-mM Tris–HCl, pH 7.5 by incubating for 30 min at 37 °C.
11. Determine ligation efficiency by running a small aliquot (1–5 µl) of 3C material on a 0.6 % (wt/vol) standard agarose gel. A successful ligation should result in a significant shift of the digested DNA upward in the gel (Fig. 3) (*see* **Note 18**).

12. Store the 3C library at −20 °C or proceed with the second round of digestion. If desired, quantitative PCR can be performed on the 3C library to either ensure the presence of chimeric ligation products or validate library quality by probing for a known genomic interaction (*see* **Note 19**).

3.5 Generation of the 4C-Seq PCR Template: Digestion and Ligation

1. Digest the 3C library overnight with the selected four-base recognition restriction enzyme using 50 U of enzyme in a 500-µl total reaction volume. Use buffers and incubation temperatures as recommended in the manufacturer's instructions.
2. Remove a small aliquot (10–20 µl) from the reaction to assess digestion efficiency on a 1.5% (wt/vol) standard agarose gel. The high molecular weight DNA observed after the first ligation should have been digested into a smear of low molecular weight fragments (usually the majority is <1000 bp) (Fig. 3). If digestion is suboptimal, consider repurifying the samples using a phenol–chloroform extraction (as described below) followed by an additional round of overnight digestion.
3. If digestion was successful, heat-inactivate the restriction enzyme by incubating at 65 °C for 20 min. If the enzyme cannot be heat-inactivated, remove the enzyme by adding 500 µl of phenol/chloroform/isoamyl alcohol. Shake vigorously and centrifuge for 15 min at 16,100 × *g*. Transfer the aqueous phase to a new tube and add 2 µl of glycogen, 50 µl of sodium acetate, and 850 µl of 100% ethanol. Mix thoroughly, place the sample at −80 °C until frozen solid, and directly centrifuge the frozen samples for 20 min at 16,100 × *g* (4 °C). Wash the pellet once with 70% ethanol, air-dry, and dissolve in 500 µl of 10-mM Tris–HCl, pH 7.5.
4. Transfer the sample to a 50-ml centrifugation tube. Add 1.4 ml 10× ligation buffer, 200 U of T4 DNA ligase, and water up to 14 ml. Incubate at 16 °C for 4 h or overnight (*see* **Note 20**).

3.6 DNA Purification, PCR Amplification, and 4C-Seq Library Preparation

1. To the 14-ml ligation sample, directly add 14-µl glycogen and a 1/10 volume of sodium acetate, mix the contents, and add 35 ml of 100% ethanol (*see* **Note 21**).
2. Place the tubes at −80 °C for 2–3 h until the liquid is frozen solid.
3. Directly centrifuge the frozen samples for 45 min at 3200 × *g* (4 °C).
4. Remove the supernatant and add 10 ml of 70% ethanol.
5. Centrifuge the mixture for 15 min at 3200 × *g* (4 °C).
6. Remove the supernatant, air-dry, and dissolve the pellet in 150 µl of 10-mM Tris–HCl, pH 7.5 by incubating for 30 min at 37 °C.

7. Purify the DNA using a spin column-based DNA purification kit (QIAquick gel or PCR purification kits work well in our hands) according to the manufacturer's recommendations. Realize that most columns used in such kits have a maximum binding capacity of 10–20 μg of DNA. Use one column when working with $1–2 \times 10^6$ cells, increase the number of columns to two ($2–5 \times 10^6$ cells) or three ($5–10 \times 10^6$ cells) as more cells are used at the start of the protocol.
8. Estimate the DNA concentration of the resulting "4C-Seq PCR template" sample using spectrophotometric (e.g., Nanodrop) or fluorometric (e.g., Qubit) measurements.
9. Store the 4C-Seq PCR template at −20 °C or proceed with inverse PCR.
10. PCR primer sets designed in Subheading 3.1 can now be tested on dilution series of the 4C-Seq PCR template. Ensure that primer pairs yield reproducible fragment smears and amplify in a linear fashion (*see* **Note 22** and Fig. 3).
11. After assessing primer pair quality, perform several PCR reactions (we generally amplify the equivalent of 500–1000-ng input DNA per bait fragment, using the Roche Expand Long Template PCR System) using PCR primers containing Illumina P5/P7 adapters as overhangs. The PCR reaction setup and program are indicated below. Ensure that the amount of input 4C-Seq PCR template used allows for a linear and reproducible PCR reaction. We generally do not exceed 200 ng of input material.

PCR reaction mix

5 μl of 1× buffer I

1 μl of 10 mM dNTPs

25-pmol forward primer

25-pmol reverse primer

0.75-μl polymerase mix (5 U/μl)

25–200 ng of 4C-Seq PCR template

H_2O up to 50 μl.

PCR program

94 °C—2 min

30 cycles: 94 °C, 15 s; [primer-specific Tm] °C, 1 min; 68 °C, 3 min

68 °C—7 min

12. Verify PCR success by analyzing a small aliquot (5–10 μl) of each reaction on a 1.5 % (wt/vol) standard agarose gel (Fig. 3).
13. Pool all successful reactions from the same bait fragment and purify the DNA using the High Pure PCR Product Purification kit or AMPure XP beads (*see* **Note 23**). Elute the columns/beads with 40 μl of elution buffer.

14. Verify the purification procedure success by running a 5-μl aliquot on a 1.5% (wt/vol) standard agarose gel and measuring sample purity using spectrophotometric measurements. We adhere to similar guidelines as those published by van de Werken et al. [19], with $A_{260/A280}$ values between 1.8 and 2.0 and $A_{260/A230}$ values >1.5 considered compatible with Illumina sequencing procedures. Also ensure that primer dimers (70–120 bp) have been successfully removed.
15. 4C-Seq libraries can be stored at −20 °C for at least 6 months and can be directly sequenced using Illumina high-throughput sequencing equipment (*see* **Note 24**).

3.7 High-Throughput Sequencing

1. Estimating the average fragment size of 4C-Seq libraries can be challenging. We quantify the average fragment size of the individual 4C-Seq libraries on an Agilent Bioanalyzer (ensure that the cartridge and kit used allow detection of DNA fragments up to 4 kb). If PCR smears show one or two dominant fragments, this size (or the average of the two) is generally a good approximation. Otherwise, a "smear analysis" quantification using the Bioanalyzer software provides a reliable size estimate.
2. Quantify 4C-Seq library DNA concentrations using a reliable library quantification strategy. We have successfully used Bioanalyzer, Qubit, and quantitative PCR (i.e., KAPA Library Quantification Kit) strategies.
3. Determine DNA molarities of the individual 4C-Seq libraries using the average size estimate and DNA concentration values. The following formula can be used: (library concentration [in ng/μl] × 10^6)/(650 D × library size [in bp]) = *x*nM 4C-Seq library.
4. Create different pools of multiple 4C-Seq libraries for Illumina sequencing by pooling equimolar amounts of individual libraries in a single tube (*see* **Note 25**).
5. Proceed with single-read sequencing as described by the manufacturer. Depending on the primer design, either a 50- or 100-bp read length is sufficient for accurate alignment of the reads (*see* **Note 10**). To enable custom de-multiplexing using the viewpoint primer at the beginning of the 3C-/4C-Seq sequence reads, do not specify the "I7_index_ID" and "index" columns in the SampleSheet.csv loaded upon starting the sequencer.

3.8 Primary Data Analysis

1. Proceed with standard sequencing data processing methods from Illumina (*bcl2fastq*) to retrieve the reads in FastQ format. After this procedure, all data will be located (by default) in the data run folder, in the Data/Intensities/BaseCalls/subfolder, organized as one gzipped FastQ file per lane.

```
> bcl2fastq --runfolder-dir /path/to/run-
folder/
```

```
> cd /path/to/runfolder/Data/Intensities/
BaseCalls; gunzip *.fastq.gz
```

2. Obtain the reads from the samples by matching the 5′-ends to the viewpoint sequence (*see* **Note 27**).

```
> demultiplex.py -i allreads.fastq -s view-
point_sequence -o sample.fastq
```

3. Next, remove the viewpoint and the restriction site from the 5′-end of the reads to keep only the interacting target sequence. The remaining sequence should be trimmed to a fixed length from the 3′-end of the read. Differences in read length between different viewpoints can introduce variation between experiments and should thus be avoided. In our experience 36 base-pair reads yield good mappability and high sensitivity to a wide size range of restriction fragments.

```
> subsection.py -i sample.fastq \
      -o infile.fastq \
      -s length_of_viewpoint_plus_restric-
tion_site \
      -e length_of_viewpoint_plus_restric-
tion_site+36
```

4. Check for the presence of primary restriction enzyme recognition sites in the reads. Restriction sites can sometimes be absent from a subset of the reads, which is likely due to unexpected endonuclease activity and/or PCR artifacts. To check for such issues, the bases comprising the restriction site are extracted from the reads, and the percentage of reads containing the expected restriction site is determined.

```
> subsection.py -i sample.fastq \
      -o sites.txt \
      -s length_of_viewpoint \
      -e length_of_viewpoint+6 \
      --table
> cut -f 2 sites.txt | sort | uniq -c | sort
-nr>sites.cnt
> grep -e recognition_sequence sites.cnt
```

   As a rule of thumb, over 90% of the reads should contain the primary restriction enzyme site.

5. Align the trimmed reads to the reference genome with *Bowtie* [14] (*see* **Note 28**).

```
> bowtie -qS -l 32 -n 2 -p 4 --best -m 1 /
path/to/genome/index \
    infile.fastq>infile.sam 2>bowtie.error.log
> samtools view -Sb infile.sam>infile.bam
> samtools sort infile.bam infile.srt
> samtools index infile.srt.bam
```

6. Convert the BAM files into BEDgraph files for visualization and quality control (*see* **Note 29**).

```
> echo 'track type=bedGraph name=$*
visibility=full'>infile.cvrg.bedgraph
> bedtools genomecov -ibam infile.srt.bam
>>infile.cvrg.bedgraph
```

3.9 Downstream Data Analysis

1. Generate a restriction map of the target genome.

```
> findSequence.py -f genome.fasta -s recog-
nition_sequence -b occurences.bed
> regionsBetween.py -I occurences.bed -s
chromsizes.txt -o regions.bed
> bedtools sort -I regions.bed>regions.srt.
bed
```

2. Count the number of reads per restriction fragment.

```
> alignCounter.py -b infile.srt.bam -r re-
gions.srt.bed -o infile.table
```

3. Convert the tables to BEDgraph files for visualization in genome browsers.

```
> echo 'track type=bedGraph name=$*
visibility=full'>infile.bedgraph
> gawk '/[^#]/{if($4>0) print $1 "\t" $2
"\t" $3 "\t" $4;}' \
       infile.table >>infile.bedgraph
```

4. (*Optional*) Convert the BEDgraph files to .tdf files for viewing in the IGV genome browser. IGV can also perform this indexing upon loading BEDgraph files.

```
> java -Xmx2g -Djava.awt.headless=true -jar
igvtools.jar \
     -f median,mean,max -z 10 infile.bedgraph
file.tdf genome-name
```

5. Verify 4C-Seq library quality by examining several important parameters. First and foremost, the vast majority (>50%) of reads should map back to the cis-chromosome, clearly clustering in close proximity (i.e., within 1 Mb) of the viewpoint. Read count percentages on the viewpoint fragment itself and the fragment directly adjacent to the location of the reading primer (the "self-circle" and "undigested" prominent signals, respectively; *see* Figs. 2 and 3) are informative for assessing crosslinking and digestion efficiencies. Van de Werken et al. [19] provide excellent information on 3C-/4C-Seq library quality control parameters.

6. Normalization, in-depth statistical analysis, and visualization of 3C-/4C-Seq data can be performed using several existing software packages. 3C/4C data analysis is usually performed using either a restriction fragment-based approach or a window-based approach. Using the restriction fragment-based

approaches, the reference genome is digested with the restriction enzyme in silico (*see* **step 1** above). These restriction fragments are then assigned a score based on the number of reads aligning to that fragment (*see* **step 2** above). In window-based approaches, sequencing read enrichments are determined per chosen genomic interval (i.e., a 20-kb window). Restriction fragment-based approaches normally yield bar and line graphs as a standard visual output (Fig. 4a, b). Window-based approaches are often represented using smoothened line graphs and "domainograms" (Fig. 4b, c). Below, we briefly describe the main 3C-/4C-Seq analysis software packages currently available.

(a) *r3Cseq* [20] is a Bioconductor package available for the R statistical software environment. The package performs data normalization, identifies statistically significant interactions (within and between experimental datasets), and provides several options for data visualization. The package uses BAM alignment files to count the number of reads per restriction fragment. These are then normalized using a reverse-cumulative fitting procedure in which the area around the viewpoint is not considered. Significant interactions are detected by comparing the observed read counts per fragment to those of a smoothened background signal. Biological replicates are combined using Fisher's combined probability test. r3Cseq creates restriction fragment bar graphs, log2 fold change comparison tracks, line plots, and domainograms.

(b) *Basic4Cseq* [21] is another Bioconductor package available in R. This software is focused on data visualization, but does provide data smoothening and a basic "reads-per-million" normalization function. Basic4Cseq visualizes the raw and normalized data per restriction fragment across the genome and creates running mean and median plots for varying window sizes. Data can be exported in either the .csv or .wig formats for downstream processing using other tools.

(c) *fourSig* [22] is a command-line tool written in Perl and R. It performs a per restriction fragment analysis, but also features a novel statistical approach to determine significant interactions with greater confidence ("interaction prioritization").

Fig. 4 (continued) frequent interactions between transcription factor-bound regions and the *MYB* promoter. Panel **c** displays 4C-Seq data as a so-called domainogram generated using r3CSeq software [20], in which an increasing window size (2–30 kb resolution) is used to identify broader regions of high interaction frequency. Two conditions (red blood cell progenitors and brain cells) were analyzed. Note how certain regions in the *Myb* locus display strong tissue-specific interactions with the *Myb* promoter. Many of the red blood cell-specific interactions contain transcription factor-bound enhancer regions, in agreement with the much higher *Myb* expression levels in these cells. Data were taken from Stadhouders et al. [5, 16]

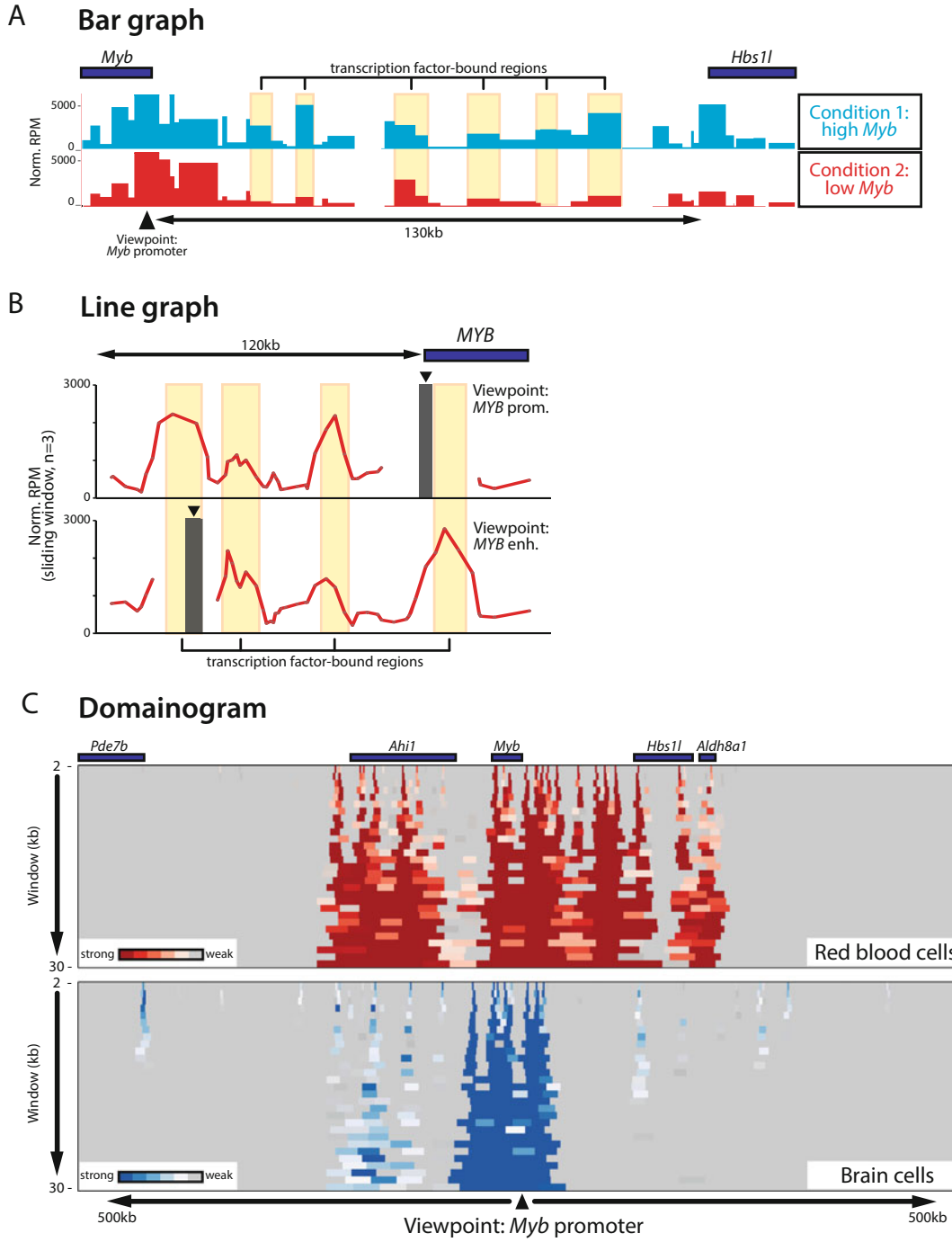

Fig. 4 Examples of 4C-Seq data visualization. 4C-Seq data can be displayed using different strategies. In the different examples shown, the mouse (panels **a** and **c**) or human (panel **b**) *Myb*/*MYB* promoter was used as a viewpoint in order to detect genomic regions in the near vicinity interacting with the gene. Panel **a** shows a standard bar graph representation, with each individual bar representing the (normalized) number of sequences retrieved from that particular restriction fragment—a measure for the interaction frequency between that fragment and the viewpoint. Two conditions (red blood cell progenitors and differentiating red blood cells) were analyzed. Note that interactions between the *Myb* promoter and transcription factor-bound intergenic regions (marked in *yellow*) are lost upon differentiation. Panel **b** depicts 4C-Seq data as a line graph, again highlighting

In terms of visualization, fourSig employs restriction fragment read count files for genome browser visualization.

(d) *4Cseqpipe* [23] software is a complete analysis package that not only performs data normalization and visualization but also includes a specific 4C-specific aligner that matches reads with the theoretical restriction fragments in the genome. The processed data is visualized in several ways, including high-resolution domainograms and genome-wide ideograms.

(e) *FourCSeq* [24] is an R package that uses a variance-stabilizing transformation and a trend-fitting approach for advanced data normalization. Statistically significant interactions above the expected background trend can be compared between experimental conditions, resulting in line graphs and log2 fold change comparison tracks that highlight significantly different interactions.

4 Notes

1. Any mammalian cell is in principle suitable for 4C-Seq analyses. To ensure equivalent formaldehyde exposure (and thus comparable crosslinking efficiencies) among individual cells and between different cell types, it is highly desirable to obtain single-cell suspensions before crosslinking. Especially when using tissue samples, the use of cell strainers and even collagenase treatment should be considered.
2. Typically, formaldehyde used in 3C (and many other biochemical assays, such as chromatin immunoprecipitation) comes from 37% formaldehyde stocks that contain 10–15% methanol as a stabilizing agent. Although suitable for 3C-based technologies, one should realize that over time these formaldehyde stocks are subjected to polymerization and oxidation. We consider the maximum shelf life of a 37% formaldehyde bottle to be 6 months after opening. Alternatively, one could use single-use ampules of (methanol-free) formaldehyde.
3. The recipe used here has not changed since the original publication of 3C by Dekker et al. [25], and it remains the most common lysis buffer for 3C-based approaches. It allows for (a partial) extraction of nuclei under mild conditions. As efficient lysis is considered important for allowing the restriction enzyme access to the chromatin, other lysis buffers as well as the use of a douncer have been employed (i.e., in Splinter et al. [10]) to increase lysis efficiency. Nevertheless, one should strive to be as gentle as possible considering that relevant chromatin co-associations originate from intact nuclei [26, 27]. In our hands, cell lysis is often poor using only the original 3C

lysis buffer. This is usually not problematic, as SDS addition (0.25–0.5 %) in the subsequent steps prior to digestion often efficiently lyses the cell membrane while keeping crosslinked nuclei intact. Nevertheless, some cell types are resistant to lysis (even in the presence of SDS), and it is therefore important to monitor cell lysis during the first phase of the procedure to obtain high digestion efficiencies.

4. As described in Subheading 3.1 (**step 2**), primary restriction enzyme choice is primarily determined by general performance under 3C conditions, sensitivity to CpG methylation, and ligation efficiencies of the resulting fragment ends. Other important parameters are recognition site distribution around the viewpoint of choice, as well compatibility with the second restriction enzyme (*see* Fig. 2 and its legend for a detailed example).
5. When diluting the nuclei as traditionally done in 3C-based protocols, large quantities of T4 DNA ligation buffer are used. We routinely produce our own ligation buffer using the manufacturer's standard recipe, which is stored at −20 °C as single-use aliquots.
6. The second restriction enzyme digestion is not performed on crosslinked chromatin under the somewhat harsh 3C conditions. Therefore, any 4-bp-recognizing restriction enzyme insensitive to CpG methylation leaving fragment ends with high ligation efficiencies can be used. Important is to ensure that the secondary restriction enzyme cuts the viewpoint fragment while creating a DNA fragment suitable for self-circularization in the second ligation (>200 bp) and inverse PCR primer design (*see* Fig. 2 and its legend for a detailed example).
7. After two rounds of digestion–ligation, 4C-Seq PCR template samples remain contaminated with impurities that could potentially affect the inverse PCR. Standard spin column-based DNA purification kits improve sample purity, although in our hands the persistence of suboptimal 260/280 absorbance ratios does not influence PCR efficiency and linearity under the conditions described in Subheading 3.6.
8. Final 4C-Seq library purification can be performed using spin column-based DNA purification kits or AMPure XP beads. Important at this step is to remove as much unused primer and primer dimer as possible, which resides in the <100–120-nt range. Substantial amounts of remaining primer dimers will negatively affect sequencing yield and quality. The Roche High Pure PCR Product Purification kit suggested by Splinter et al. [10] performs well in separating informative PCR product from primers and primer dimers.
9. 4C-Seq libraries have a broad size range (120 bp to >3 kb, *see* Fig. 3), which should be taken into account when choosing Agilent Bioanalyzer cartridges for quantification.

10. Standard primer design guidelines (for, i.e., length, %GC, annealing temperature, and secondary structures) can be followed. Critical is to position the reading primer—located near the primary restriction enzyme site—as close to the enzyme recognition site as possible, preferably on top (Fig. 2). As single-read sequencing commences from this side only, any extra sequence in between the primer and the primary restriction site will lower the number of nucleotides available for mapping sequences stemming from unknown interaction fragments to the genome (*see* Fig. 2 and Subheading 3.8). The design of the primer near the secondary restriction site allows for more flexibility (within ±100 bp of the enzyme recognition site, *see* Fig. 2), as sequencing is not initiated from this site. Ensure that PCR primers used to construct the final 4C-Seq libraries contain the appropriate P5 and P7 Illumina adapters as overhang: place the P5/P7 sequence directly upstream of the annealing part. Short 3–6-nt barcodes can be placed in between the P5 adapter, and annealing part of the reading primer if multiplexing of libraries generated with the same viewpoint PCR primers is desired. Adapter sequences: P5: 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′;P7: 5′-CAAGCAGAAGACGGCATACGA-3′.
11. Formaldehyde crosslinking efficiency is influenced by temperature. Ensure that solutions in which cells are crosslinked are at room temperature.
12. Published 3C-based protocols typically employ a 1–2% range of final formaldehyde concentration when crosslinking mammalian cells. Although only few systematic comparisons have been published (especially when using mammalian cells), variations in % formaldehyde within this range appear to have little influence on experimental outcome. It should be noted that higher formaldehyde concentrations can reduce digestion efficiencies [28].
13. Using more than 10×10^6 cells as starting material can promote the formation of nuclear aggregates. This phenomenon is very cell type specific and seems to correlate with cell size. Some aggregation is normal and not detrimental, but excessive clumping can severely reduce digestion efficiencies and should be avoided. In our hands, the best solution to reduce aggregation is to start with lower numbers of cells.
14. Increasing Triton X-100 concentrations (as suggested by Splinter et al. [10] and van De Werken et al. [19]) might augment SDS quenching and further improve digestion efficiency.
15. Restriction enzymes are known to vary with respect to their ability to remain active during prolonged incubation times. In case a primary restriction enzyme is used with reported low survivability, multiple separate and/or extra enzyme additions can improve digestion efficiency.

16. Digestion efficiencies can also be quantified more accurately using quantitative PCR analysis (qPCR) employing primer sets that span restriction sites. One should strive for >70% overall digestion efficiencies.
17. This protocol still contains the traditional extreme dilution of the first 3C ligation initially implemented to favor intramolecular ligation and to minimize random ligation events. Although not detrimental, the benefit of ligating crosslinked chromatin under such dilute conditions is now considered obsolete [26, 27]. Therefore, primary 3C ligations can also be performed in smaller volumes (e.g., as described in Rao et al. [27]).
18. Bear in mind that average DNA fragment sizes after primary ligation differ depending on whether a 6-bp- or 4-bp-recognizing restriction enzyme was used: ligations after digestion with the latter type of enzyme tend to produce a less sharp high molecular weight band.
19. At this point, the 3C library can be subjected to quantitative interaction analysis using qPCR. Both SYBR Green and TaqMan probe approaches can be used to quantify interactions in a one-versus-one manner. This can be extremely useful when verifying key interactions identified using more high-throughput methodologies such as 3C-/4C-Seq. Important considerations for such 3C-(q)PCR assays have been described elsewhere [29].
20. In contrast to the primary ligation step (also *see* **Note 17**), the second ligation (specific for 3C-/4C-Seq protocols) needs to be performed under extremely diluted conditions to strictly promote self-circularization of individual fragments and to prevent random ligations between DNA molecules in the solution.
21. Traditionally, 3C-based protocols include several phenol/chloroform purifications steps. In the original 4C-/4C-Seq protocols [11, 30], ligation products at this stage were first purified by a phenol/chloroform/isoamyl alcohol extraction, immediately followed by ethanol precipitation and spin column-based purification. To prevent additional loss of material when working with low numbers of cells, we now omit the phenol/chloroform extraction and immediately continue with the ethanol precipitation.
22. A standard setup for testing 4C-Seq PCR primers involves running a duplicate series of PCR reactions (using the reaction setup and program described in **step 11** of Subheading 3.6) using 25, 50, and 100 ng of 4C-Seq PCR library input DNA. DNA smears are visualized on a 1.5% standard agarose gel to determine reproducibility and linearity. Another hallmark of a successful 4C-Seq PCR is the appearance of two "prominent" bands representing self-circularization of the viewpoint fragment and ligation of the viewpoint fragment to

the adjacent fragment when digestion of the primary restriction site was not achieved (Figs. 2 and 3). Both these events are much more prominent than other interactions, hence their abundance in 3C-/4C-Seq libraries. The size of the expected prominent bands can be easily estimated (Fig. 2). Note that these fragments can be very short and therefore not compatible with self-circularization, or they can simply not be visualized on a 1.5 % agarose gel.

23. One column is sufficient when using the High Pure PCR Product Purification kit.
24. The P5/P7 single-read Illumina sequencing adapters used in this protocol (*see* **Note 10** and Stadhouders et al. [11]) are not compatible with MiSeq and NextSeq sequencing platforms and should be sequenced on HiSeq or GA instruments. If sequencing on MiSeq/NextSeq platforms is desired, the P5/P7 sequences described here should be swapped for adapter sequences appropriate for these instruments.
25. Important to realize is that Illumina sequencing instruments use the first four bases to recognize DNA clusters on the flowcell. Especially when reading these first four bases, nucleotide complexity needs to be high—too little variation and base calling will be compromised—leading to reduced sequence yields. As 3C-/4C-Seq libraries consist of amplicons all starting with the same sequence (the viewpoint-specific primer), pools of different (>6) 4C-Seq libraries (either using different viewpoint PCR primers or different barcodes placed upstream of the viewpoint-specific reading primer sequence) have to be sequenced together in a single lane to create enough sequence diversity. Alternatively, one could combine 4C-Seq libraries with other samples not suffering from these diversity issues (e.g., RNA-Seq, ChIP-Seq samples, or a PhiX sequencing control sample).
26. All analysis software has been tested on Red Hat Enterprise Linux Server release 6.4 (Santiago), but should work on any Linux distribution.
27. In most situations, reads can be assigned to specific samples based on the first ten bases. Using more bases may cause data yield to suffer as sequence errors accumulate across the reads. Using fewer bases (down to six) is possible when the number of viewpoints is small and their sequences differ at the 5′-end.
28. Bowtie does not align reads with insertions or deletions, but this is generally not an issue in 3C-/4C-Seq data analysis. If insertions and deletions need to be taken into account, BWA [31] can be used.
29. BAM files can in principle be viewed in genome browsers directly, but the large numbers of reads present near the viewpoint location prevent efficient visualization on most personal

computers. BEDgraph files only contain the read–depth per genome position and thus allow 4C-Seq data to be rapidly viewed on any standard personal computer. We regularly use the standalone application IGV (*see* software list) or the web-based UCSC genome browser (https://genome.ucsc.edu/), although any genome viewer can be used.

Acknowledgments

The authors would like to thank the members of the Soler and Van IJcken laboratories for insightful discussions. R.S. is supported by an EMBO Long-term Fellowship (ALTF 1201-2014) and a Marie Curie Individual Fellowship (H2020-MSCA-IF-2014, proposal #654933). E.S. is supported by grants from the FSER (Schlumberger Foundation for Education and Research), ARC Foundation ("project ARC" n°SFI20121205625), and the Atip-Avenir program (Plan Cancer).

References

1. de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26(1):11–24. doi:10.1101/gad.179804.111
2. Pombo A, Dillon N (2015) Three-dimensional genome architecture: players and mechanisms. Nat Rev Mol Cell Biol 16(4):245–257. doi:10.1038/nrm3965
3. de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502(7472):499–506. doi:10.1038/nature12753
4. Visser M, Kayser M, Palstra RJ (2012) HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 22(3):446–455. doi:10.1101/gr.128652.111
5. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S, Grosveld F, Thein SL, Soler E (2014) HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J Clin Invest 124(4):1699–1710. doi:10.1172/JCI71520
6. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nobrega MA (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507(7492):371–375. doi:10.1038/nature13138
7. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, Beckwith CA, Chan JA, Hills A, Davis M, Yao K, Kehoe SM, Lenz HJ, Haiman CA, Yan C, Henderson BE, Frenkel B, Barretina J, Bass A, Tabernero J, Baselga J, Regan MM, Manak JR, Shivdasani R, Coetzee GA, Freedman ML (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet 41(8):882–884. doi:10.1038/ng.403
8. Cowper-Sal lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, Moore JH, Lupien M (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet 44(11):1191–1198. doi:10.1038/ng.2416
9. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L, Sabo PJ, Vierstra J, Voit RA, Yuan GC, Porteus MH, Stamatoyannopoulos JA, Lettre G, Orkin SH (2013) An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342(6155):253–257. doi:10.1126/science.1242088
10. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58(3):221–230. doi:10.1016/j.ymeth.2012.04.009
11. Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E (2013) Multiplexed chromosome conformation capture sequencing for rapid genome-scale

high-resolution detection of long-range chromatin interactions. Nat Protoc 8(3):509–524. doi:10.1038/nprot.2013.018

12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. doi:10.1093/bioinformatics/btp352
13. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842. doi:10.1093/bioinformatics/btq033
14. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25. doi:10.1186/gb-2009-10-3-r25
15. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26. doi:10.1038/nbt.1754
16. Stadhouders R, Thongjuea S, Andrieu-Soler C, Palstra RJ, Bryne JC, van den Heuvel A, Stevens M, de Boer E, Kockx C, van der Sloot A, van den Hout M, van Ijcken W, Eick D, Lenhard B, Grosveld F, Soler E (2012) Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J 31(4):986–999. doi:10.1038/emboj.2011.450
17. Ribeiro de Almeida C, Stadhouders R, Thongjuea S, Soler E, Hendriks RW (2012) DNA-binding factor CTCF and long-range gene interactions in V(D)J recombination and oncogene activation. Blood 119(26):6209–6218. doi:10.1182/blood-2012-03-402586
18. Stadhouders R, de Bruijn MJ, Rother MB, Yuvaraj S, Ribeiro de Almeida C, Kolovos P, Van Zelm MC, van Ijcken W, Grosveld F, Soler E, Hendriks RW (2014) Pre-B cell receptor signaling induces immunoglobulin kappa locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. PLoS Biol 12(2), e1001791. doi:10.1371/journal.pbio.1001791
19. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112. doi:10.1016/B978-0-12-391938-0.00004-5
20. Thongjuea S, Stadhouders R, Grosveld FG, Soler E, Lenhard B (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res 41(13), e132. doi:10.1093/nar/gkt373
21. Walter C, Schuetzmann D, Rosenbauer F, Dugas M (2014) Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. Bioinformatics 30(22):3268–3269. doi:10.1093/bioinformatics/btu497
22. Williams RL Jr, Starmer J, Mugford JW, Calabrese JM, Mieczkowski P, Yee D, Magnuson T (2014) FourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res 42(8):e68. doi:10.1093/nar/gku156
23. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A, de Laat W (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9(10):969–972. doi:10.1038/nmeth.2173
24. Klein FA, Pakozdi T, Anders S, Ghavi-Helm Y, Furlong EE, Huber W (2015) FourCSeq: analysis of 4C sequencing data. Bioinformatics. doi:10.1093/bioinformatics/btv335
25. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. doi:10.1126/science.1067799
26. Gavrilov AA, Gushchanskaya ES, Strelkova O, Zhironkina O, Kireev II, Iarovaia OV, Razin SV (2013) Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. Nucleic Acids Res 41(6):3563–3575. doi:10.1093/nar/gkt067
27. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. doi:10.1016/j.cell.2014.11.021
28. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W (2003) The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 35(2):190–194. doi:10.1038/ng1244
29. Hagege H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forne T (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). Nat Protoc 2(7):1722–1733. doi:10.1038/nprot.2007.243
30. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38(11):1348–1354. doi:10.1038/ng1896
31. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. doi:10.1093/bioinformatics/btp324

Chapter 16

Using an Inducible CRISPR-dCas9-KRAB Effector System to Dissect Transcriptional Regulation in Human Embryonic Stem Cells

Krishna Mohan Parsi, Erica Hennessy, Nicola Kearns, and René Maehr

Abstract

CRISPR-Cas9 effector systems have wide applications for the stem cell and regenerative medicine field. The ability to dissect the functional gene regulatory networks in pluripotency and potentially in differentiation intermediates of all three germ layers makes this a valuable tool for the stem cell community. Catalytically inactive Cas9 fused to transcriptional/chromatin effector domains allows for silencing or activation of a genomic region of interest. Here, we describe the application of an inducible, RNA-guided, nuclease-deficient (d) Cas9-KRAB system (adapted from *Streptococcus pyogenes*) to silence target gene expression in human embryonic stem cells, via KRAB repression at the promoter region. This chapter outlines a detailed protocol for generation of a stable human embryonic stem cell line containing both Sp-dCas9-KRAB and sgRNA, followed by inducible expression of Sp-dCas9-KRAB to analyze functional effects of dCas9-KRAB at target loci in human embryonic stem cells.

Key words Human ES cells, CRISPR-Cas9, dCas9-KRAB, Transcription, Gene regulation, Gene silencing and activation

1 Introduction

Embryonic stem (ES) cells are characterized by an unlimited self-renewal capacity and the ability to differentiate into cell types of all three germ layers. These unique characteristics of ES cells provide a valuable in vitro model to understand the early aspects of human development. The in vitro expansion and differentiation potential of human ES cells can provide an unlimited source of various developmental intermediates that can be used for drug screening and possibly regenerative medicine purposes. Currently, a major bottleneck to regenerative medicine applications of human ES cells is identifying how to direct human ES cells to the desired differentiated cell type of interest, with high efficiency and purity. Hence, the identification and characterization of cell-type-specific regulatory elements and key factors that maintain regulatory networks has

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_16,

become an intense topic of investigation in the field of developmental biology and regenerative medicine. In order to better understand the cellular state-specific gene regulatory network, efficient gene silencing and/or activation methods are highly desirable.

The newly adapted microbial clustered regularly interspaced short palindromic repeat (CRISPR) system has been shown to efficiently target DNA regions in an RNA-guided manner allowing dissection of gene function in a given cell type [1, 2]. The CRISPR system is an efficient adaptive immune response mechanism used by many prokaryotes to degrade foreign DNA molecules [3, 4]. Modified versions of Type II CRISPR-Cas9 systems have been applied efficiently in eukaryotes to edit specific endogenous genomic regions [5]. They require two main components, Cas9 nuclease and guide RNA (single chimeric guide RNA—sgRNA), which has a fused version of crRNA and tracrRNA [1, 6]. When these two components are provided in a given cell, the expressed sgRNA directs Cas9 nuclease to endogenous complementary DNA sequences where the Cas9 induces a dsDNA break [6]. Type II CRISPR systems identified in bacterial species such as *Streptococcus pyogenes* (Sp), *Neisseria meningitidis* (Nm), *Francisella novicida* (Fn), *Streptococcus thermophilus 1* (St1), *Treponema denticola* (Td), and *Staphylococcus aureus* (Sa) were adapted for application in eukaryotes due to their single Cas9 effector protein with minimal regulatory sequence requirements for target recognition and cleavage [7–9]. A short genomic sequence adjacent to the protospacers called protospacer adjacent motif (PAM) is crucial for Cas9 targeting, and to introduce dsDNA breaks at the target DNA site [10], with each Cas9 ortholog requiring different PAM sequences for their Cas9 nuclease function. The length and complexity of these PAM sequences impact the frequency that these sequences are found in the target genome and, as such, the number of regions that may be targeted by each Cas9 ortholog. The Sp-Cas9 requires a short NGG PAM sequence leading to its popularity as a versatile system [11]. The Sp-Cas9 system also functions, but to a lesser extent, with a NAG PAM sequence [12].

CRISPR-Cas9-induced deletion, insertion, or modification to the target endogenous DNA locus can produce irreversible changes to DNA and gene expression. In order to modulate gene expression without altering the target DNA sequences, RNA-guided programmable catalytically inactive Cas9 (dCas9) is used, which can be fused with various chromatin/transcriptional effectors [12–20]. dCas9 fused with transcriptional repressor domains (such as Kruppel-associated box (KRAB) repressor or SID domains) silences the target loci—termed CRISPR transcriptional interference (CRISPRi) [12], whereas dCas9 fused with transcriptional activation domains (such as VP16, VP64, p65AD domains) activates the target loci—termed CRISPR transcriptional activation (CRISPRa) [17]. Targeting dCas9 alone to the gene promoter can also repress

transcription due to steric hindrance of the transcriptional complex [15, 18]. In addition to transcriptional effector domains, chromatin modifiers such as LSD1 and p300 fused with dCas9 protein silence and activate the target genomic loci, respectively [19, 20].

Unlike small RNA-mediated loss-of-function screens, which involve posttranscriptional silencing of RNA in the cytoplasm, the CRISPR system has shown to be more effective in modulating RNA levels by influencing the transcriptional status of the target [21, 22]. The CRISPR system also allows a wider coverage of the genome, making it possible to use a single platform to probe functional effects of both protein coding and noncoding regions [19, 20]. Using dCas9 effectors for genome-wide applications has great scope for dissecting cell fate-determining transcriptional networks. In the future, this system may be utilized to dissect transcriptional regulatory networks of human ES cell differentiation products to investigate human development and disease states.

Here, we describe a detailed protocol to manipulate gene expression in human ES cells using an inducible Sp-dCas9-KRAB system. We outline the preparation of human ES cells for manipulation, generation of an inducible stable dCas9-KRAB ES cell line, and delivery of sgRNAs and doxycycline-inducible expression of Sp-dCas9-KRAB to permit functional characterization of genomic targets (experimental workflow outlined in Fig. 1). While this

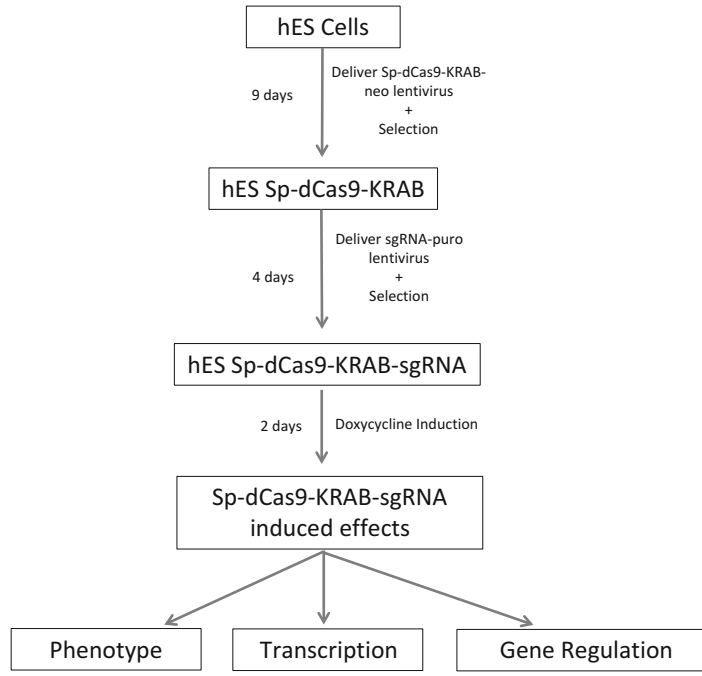

Fig. 1 Overview of an experimental workflow for generation of Sp-dCas9-KRAB and sgRNA expression in human ES cells

protocol outlines the use of a KRAB effector targeted to a gene promoter region, it can easily be adapted by substituting alternative effector domains. Additionally, other genomic sites including coding and noncoding gene regulatory regions can be targeted by manipulation of the sgRNA sequence.

2 Materials

2.1 Cell Culture Reagents

1. Cell line: H1 human ES cells (WiCell).
2. mTeSR1 Basal medium (Stem cell Technologies) complete with mTeSR 5× supplement.
3. Matrigel ® hESC qualified matrix.
4. Phosphate-buffered saline (PBS), 1×.
5. Dispase in Hank's balanced salt solution, 5 U/mL—The working stock of dispase (1 U/mL) is prepared by dilution of 1 volume of dispase with 4 volumes of DMEM/F12 media. Diluted dispase can be stored at 4 °C for a month.
6. DMEM/F12 (1:1), containing l-glutamine and 2.438 g/L sodium bicarbonate, 1×.
7. TrypLE express.
8. RHO/ROCK pathway inhibitor—Y-27632 (dihydrochloride)—Stock concentrations of 10 mM are prepared in DMSO, stored at −20 °C, and protected from light. Working concentrations of 10 μM are prepared fresh as required.
9. Tissue culture dishes, 100 × 20 mm.
10. Cell lifter, polyethylene.
11. Costar 6-well plates, ultralow attachment surface, polystyrene.
12. Geneticin, 50 mg/mL.
13. Puromycin, 10 mg/mL.
14. Doxycycline, 2 mg/mL.
15. 15 mL falcon tubes.
16. Corning 6-well plates
17. Dimethyl sulfoxide Hybri-Max.

2.2 Plasmids

1. dCas9-effector plasmid
   pHAGE TRE-dCas9-KRAB (Addgene).
2. sgRNA plasmid
   pLenti Sp BsmBI sgRNA Puro (Addgene) plasmid (modified to contain user-specific target sgRNA).

2.3 Molecular Biology Reagents

1. Rat anti-hemagglutinin (HA) antibody (3F10) (Roche).
2. Alexa Fluor 488 donkey anti-rat secondary antibody (Invitrogen) or similar.
3. Hoechst 33342, 10 mg/mL.
4. Target-specific qPCR primers and/or antibody.
5. Trizol.
6. Molecular biology grade water.
7. Superscript III first-strand synthesis kit.
8. Formalin solution, 10%.
9. Donkey serum.
10. Tween 20.
11. Eppendorf tubes, 1.5 mL.
12. SYBR-FAST universal qPCR kit.

3 Method

3.1 Maintenance of H1 Human ES Cell Cultures

Undifferentiated H1 cells are maintained and cultured under feeder-free conditions.

1. Tissue culture-grade dishes (corning 100 × 20 mm) are coated with Matrigel® hESC qualified matrix (diluted in DMEM/F12 media according to manufacturer's specifications) at 37 °C for 1 h (*see* **Note 1**).
2. Thaw cryopreserved H1 human ES cells (approx. 2×10^6 cells/vial) in a 37 °C water bath until only a small ice crystal remains; slowly add 9 mL of mTeSR media to dilute the toxic DMSO concentration.
3. Centrifuge resuspended cells at 200 × *g* for 5 min at room temperature.
4. Discard the media without disturbing the pellet.
5. Gently resuspend the pellet in 10 mL of mTeSR media with 10 μL of 10 mM Y-27632 stock (final concentration 10 μM). Add cell suspension to Matrigel matrix-coated culture plates and incubate at 37 °C and 5% CO_2 (*see* **Note 2**).
6. Feed the cells with fresh mTeSR media (without Y-27632) daily.
7. Human ES cells are passaged every 3–4 days for maintenance, to prevent over confluence or limit spontaneous differentiation. To passage cells:
   (a) Remove media and wash once with 1× PBS buffer.
   (b) Add 2 mL of 1 U/mL dispase and incubate at 37 °C for 7 min (*see* **Note 3**).

(c) Remove dispase and discard. Rinse dish twice with 10 mL of 1× PBS taking care not to remove any human ES colonies.

(d) Add 8 mL of fresh mTeSR media to the cells and gently scrape compact colonies using cell lifter.

(e) Pipette gently for 3–4 times to break the larger colonies into small uniform colonies (see Note 2).

(f) Transfer into new plates at a ratio of 1:4.

3.2 Generation of Doxycycline-Inducible Sp-dCas9-KRAB Effector Stable Line

Standard procedures are used to generate third-generation lentiviral particles from pHAGE-TRE-dCas9-KRAB and are not covered in detail here but follow published methods [16]. Standard viral handling procedures are followed during the transduction procedure and for 48 h afterwards. Successfully transduced cells containing Sp-dCas9-KRAB can be selected using Geneticin and expanded to generate stable cell lines.

1. Take one 100 mm H1 human ES cell culture plate and wash once with 1× PBS.
2. Add 2 mL of TrypLE express solution to the cells and incubate at 37 °C for 7 min to dissociate human ES colonies into single cells (*see* **Note 3**).
3. Add 8 mL of mTeSR media and pipette up and down to ensure all colonies have dissociated from the culture dish, and a uniform single cell suspension is achieved. Collect cell suspension in a 15 mL falcon tube.
4. Centrifuge cells at 200 × *g* for 5 min at room temperature.
5. Discard supernatant and resuspend cells in 5 mL of mTeSR media. Perform a cell count to determine the total number of cells per mL.
6. Plate 1×10^6 human ES cells in 1 mL of mTeSR media in one well of an ultralow attachment 6-well plate. Add Y-27632 to achieve a final concentration of 10 μM.
7. Add dropwise 1 mL of Sp-dCas9-KRAB lentiviral particles (approximate titer of 10^5 viral particles/mL) to **step 6** human ES cells (*see* **Note 4**).
8. Incubate the plate at 37 °C for 3 h. Agitate every 30 min to prevent human ES cells from settling. During this stage, human ES cells will remain in suspension, maximizing exposure to the lentiviral particles.
9. After 3 h, bring the volume up to 4 mL in mTeSR media containing 10 μM Y-27632, and distribute 2 mL per well into 2 wells of a Matrigel matrix-coated 6-well plate (*see* **Notes 1** and **2**).
10. After 24 h, media is carefully discarded and the cells are fed with fresh mTeSR media (without Y-27632).

11. 48 h after transduction, human ES cells are changed with fresh mTeSR media containing 50 μg/mL Geneticin selection reagent, to select for cells that have been successfully transduced with the Sp-dCas9-KRAB lentiviral particles.
12. Feed transduced human ES cells daily with fresh mTeSR media containing 50 μg/mL Geneticin selection reagent until non-transduced human ES control cells have died (usually 6 days).
13. To confirm the expression of Sp-dCas9-KRAB effector in stable human ES lines, cells are cultured in mTeSR media with 2 μg/mL of doxycycline and 50 μg/mL of Geneticin selection reagents for 48 h. Cells without doxycycline exposure are used as a negative control.
14. At 48 h, cells are analyzed by immunofluorescence for the HA-epitope tag to determine Sp-dCas9-KRAB expression. All steps are performed on a slowly rocking platform.
    (a) Wash cells with 1× PBS and fix with 10% formalin solution for 20 min at room temperature.
    (b) Wash cells once with 1× PBS and incubate cells in blocking buffer (5% donkey serum in PBS 0.2 % Triton X-100) for 45 min at room temperature.
    (c) After blocking, cells are incubated with rat anti-HA antibody (1:1000) in blocking buffer overnight at 4 °C.
    (d) Wash cells three times, 5 min each in PBS 0.2 % Triton X-100 buffer at room temperature.
    (e) Add secondary antibody (Alexa Fluor 488 donkey anti-rat secondary, 1:300 dilution in blocking buffer) and incubate at room temperature for 2 h.
    (f) Wash cells with 1× PBS 0.2 % Triton X-100 three times, 5 min each, at room temperature.
    (g) Stain with 4 μg/mL Hoechst stain for 5–10 min at room temperature.
    (h) Wash once with 1× PBS and view using a fluorescence microscope.
    (i) Doxycycline-treated cells should display positive staining for HA-epitope tag, indicative of Sp-dCas9-KRAB expression. Cells not treated with doxycycline should be negative or show minimal HA staining; these wells can be used for normalization of background antibody staining.

3.3 Delivery of sgRNA Plasmids into Sp-dCas9-KRAB Effector Human ES Cell Line

The sgRNA target sequence is user defined and dependent on the individual experimental aims of the user (*see* **Note 5**). Annealed sgRNA oligomers are ligated into pLenti SpBsmBI sgRNA Puro, and lentiviral particles are prepared according to standard protocols [16]. Viral handling procedures are followed during the

transduction procedure and for 48 h afterwards. Successfully transduced cells are selected using puromycin, allowing for isolation of cells containing both the Sp-dCas9-KRAB and sgRNA constructs.

1. Take one 100 mm H1 Sp-dCas9-KRAB effector human ES cell culture plate and wash once with 1× PBS.
2. Add 2 mL of TrypLE express solution to the cells and incubate at 37 °C for 7 min to dissociate human ES colonies into single cells (*see* **Note 3**).
3. Add 8 mL of mTeSR media to dilute the TrypLE action.
4. Pipette up and down to dissociate colonies into single cells and collect in 15 mL falcon tube.
5. Pellet cells at 200 × *g* for 5 min at room temperature.
6. Resuspend cells in 5 mL of mTeSR media and count the total number of cells per mL.
7. Plate 1×10^6 human ES cells in 1 mL of mTeSR media (containing 10 μM Y-27632) in an ultralow attachment 6-well plate.
8. Add dropwise 1 mL of sgRNA lentiviral particles (approximate titer of 10^6 viral particles/mL, MOI ≤ 1) to **step 7** human ES single cells (*see* **Note 4**).
9. Incubate the plate at 37 °C for 3 h in the incubator. Agitate every 30 min to prevent human ES cells from settling.
10. After the incubation, lentiviral transduced human ES cells are plated onto a Matrigel® hESC qualified matrix-coated 100 mm tissue culture dish, bringing the volume up to 10 mL with mTeSR media containing 10 μM Y-27632 (*see* **Note 2**).
11. After 24 h, virus containing mTeSR media is carefully discarded and cells are fed with fresh mTeSR media (without Y-27632).
12. The next day, add fresh mTeSR media containing 1 μg/mL of puromycin and 50 μg/mL of geneticin selection reagent to transduced human ES effector cells. Puromycin allows the selection of cells that have been successfully transduced with sgRNA lentiviral particles. Cells resistant to both puromycin and geneticin selection will contain both the Sp-dCas9-KRAB and sgRNA constructs.

3.4 Induction of Sp-dCas9-KRAB Expression

Cells are treated with doxycycline (2 μg/mL) to induce expression of the Sp-dCas9-KRAB 48 h before the initiation of the functional characterization experiment described in Subheading 3.5 (or other experiments discussed in **Note 7**). Cells not treated with doxycycline can be used as a control.

3.5 Functional Characterization of KRAB-Mediated Repression of Genomic Targets

The functional effect of sgRNA-mediated recruitment of dCas9-KRAB to a target locus can be assayed using a variety of techniques, depending on the genomic region being targeted and the downstream applications of interest for the user (*see* **Note 7**). As an example we describe an experiment where a dCas9-KRAB is targeted to a promoter region of a gene of interest (outlined in Fig. 2). Functional effects can be determined by analyzing expression of downstream target gene transcription and protein levels using qPCR and immunofluorescence, respectively. Phenotypic effects on human ES cell pluripotency and differentiation capacity can also be addressed by immunofluorescence staining and qPCR analysis of key pluripotency markers (e.g., Oct4, nanog).

1. Take one 100 mm confluent plate of sgRNA expressing Sp-dCas9-KRAB human ES cell line and wash once with 1× PBS.
2. Add 2 mL of 1 U/mL dispase and incubate at 37 °C for 7 min (*see* **Note 3**).

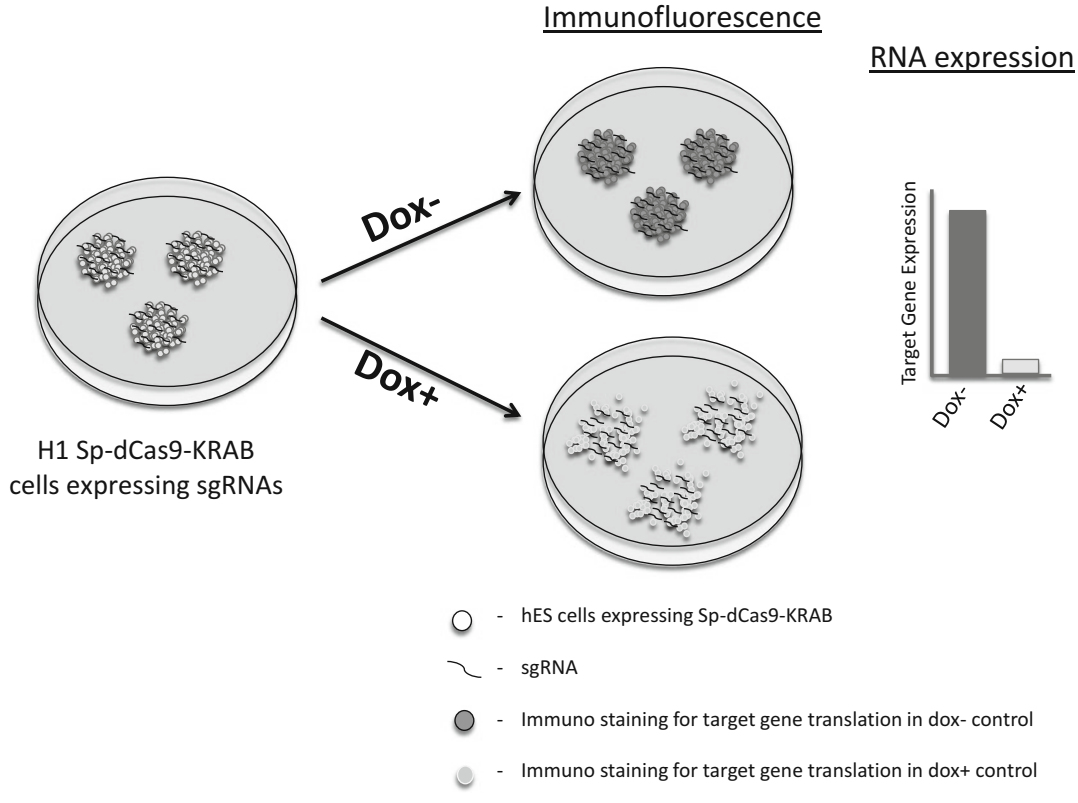

Fig. 2 Schematic representation of experimental setup for analyzing transcriptional effects of sgRNA recruitment of inducible Sp-dCas9-KRAB to a defined promoter region in human ES stable cell lines—undifferentiated human ES Sp-dCas9-KRAB cells expressing sgRNA are represented as *black circles*. Upon doxycycline induction, sgRNA function is assessed by immunofluorescence and RNA quantification of the target gene. Doxycycline-induced cells show a reduction or no expression (*light gray*) of the target gene compared to doxycycline absent cells (normal expression—*dark gray*)

3. Wash cells once in 10 mL of 1× PBS.
4. Add 8 mL of mTeSR media to the cells.
5. Gently scrape compact colonies using a cell lifter.
6. Pipette gently 3–4 times to break the larger colonies into small uniform colonies (*see* **Note 2**).
7. Distribute 1/24th of the cell suspension per well of a 6-well plate. Half the wells are labeled as doxycycline negative control (Dox–), and the other half are labeled as doxycycline positive (Dox+). For each condition, one well can be used for IF analysis and another for RNA quantification of target gene expression (i.e., the gene under the control of the promoter region targeted by the sgRNA). Similarly, plate cells for antibody negative control, no guide RNA control and negative control guide RNA (*see* **Note 6**).
8. Add 2 μg/mL of doxycycline to Dox+ cells and incubate at 37 °C.
9. After 24 h, feed the cells with fresh mTeSR media and the Dox+ wells are treated with 2 μg/mL of doxycycline.
10. Optional: Confirm dCas9 expression by immunofluorescence. Harvest one well from each condition for immunofluorescence analysis for anti-HA antibody following the protocol described in Subheading 3.2, **step 14**.
11. At the appropriate experimental timing determined by the investigator, two wells from each condition are analyzed, one for RNA quantification and the other for protein expression by immunofluorescence. RNA is extracted using Trizol reagent following the manufacturer's specifications. cDNA is synthesized from 1 μg of RNA using superscript III first-strand synthesis kit as described in the kit protocol. Real-time qPCR for target quantification is performed with SYBR-FAST reagent. Additional transcripts may be analyzed, depending on the end users' interests (*see* **Note 7**). Effect of KRAB-mediated repression at the target promoter can be quantified by differential expression of the Dox+ and Dox– cells or in comparison to other controls outlined in **Note 6**.

4 Notes

1. Thaw Matrigel® hESC qualified matrix on ice and dilute in cold DMEM/F12 media to avoid premature gelling. Protein concentration of Matrigel matrix displays lot-to-lot variation; therefore, the volume required to coat plates is lot dependent. Follow manufacturer's instructions for the lot-specific dilution required. Matrigel matrix-coated plates can be stored at 4 °C for 1 week.

2. Handle human ES cells gently; avoid excess and vigorous pipetting or breaking colonies into single cells. Maintaining medium-sized colonies improves survival after freezing. Addition of Y-27632 improves survival of cryopreserved or TrypLE express dissociated single human ES cells [23]; it inhibits dissociation-induced apoptosis of single human ES cells and also improves colony formation efficiency.
3. When passaging human ES cells, dispase is used for a maintenance split, and cells are maintained as clusters. TrypLE express is only used to achieve a single cell suspension for the purposes of transduction experiments.
4. Transduction efficiencies may vary widely between cell types. It is important to test the viral titer on your cell line of interest. For difficult-to-transduce cell types, viral supernatant may need to be concentrated by ultracentrifugation before use. It is important to titer Sp-dCas9-KRAB and sgRNA lentiviral particles (to achieve an MOI less than 1) minimizing the number of genomic integrations per cell, which would help to reduce any off-target effects [24–26].
5. When using dCas9 systems from *Streptococcus pyogenes* (Sp) or *Neisseria meningitidis* (Nm), ensure that the sgRNA constructs are designed for the correct species, as PAM target sequences which are critical for Cas9 function vary between species. A number of web-based analysis tools are available for aiding the design of sgRNAs (http://www.broadinstitute.org/rnai/public/analysis-tools/sgRNA-design and https://www.addgene.org/crispr/reference/#gRNA). To achieve an efficient CRISPRi with dCas9-KRAB, sgRNAs can be designed within a window of −200 bp to +300 bp of target transcriptional start site (TSS) [16, 21]. Targeting downstream to the TSS may yield stronger repression due to KRAB-mediated silencing and dCas9-effector-mediated steric hindrance of transcriptional complex [15]. We recommend testing more than two sgRNAs for each locus to be targeted.
6. To examine the role of a transcriptional regulatory element using the Sp-dCas9-KRAB system, a number of control conditions need to be incorporated.
   (a) No sgRNA transduction control: Control cells transduced with sgRNA backbone plasmid serve as a no sgRNA transduction control.
   (b) Negative control sgRNAs: sgRNAs can be designed to unrelated genomic regions not involved in any key features of human ES cells (e.g., mature differentiation markers). Additional negative control sgRNAs that do not target any human genomic region, e.g., sgRNA targeting the CAG (CMV-IE, chicken actin, rabbit beta-globin) sequence or a

scrambled sgRNA, can also be used as a negative controls [24]. These negative control sgRNAs ensure that any phenotypic effect of the sgRNA of interest is due to the sgRNA-specific targeting to the site of interest and not solely as a consequence of viral transduction and subsequent selection.

(c) Antibiotic negative control: The functional characterization experiment can be performed in the absence of selection reagent to rule out any possible influence of antibiotics on the final results.

(d) Multiple guide control: To verify an on-target-specific phenotype, use multiple sgRNAs targeted to the same genomic region of interest.

7. Experimental timing will vary depending on the endpoints of interest to the investigator. If targeting key pluripotency-related transcription factors such as Oct4, endpoints of interest (in addition to directly measuring RNA and protein level of the targeted gene) may include RNA sequencing to identify genes regulated by the target gene, FACS analysis to determine effect on pluripotency cell-surface marker expression, or teratoma assays as an in vivo functional measure of pluripotency.

Acknowledgments

We thank Scot Wolfe and the members of the Maehr Lab for the helpful discussion and technical advice. R. M. is supported by The Leona M. and Harry B. Helmsley Charitable Trust (2012PG-T1D026 and 2015PG-T1D035), the Charles H. Hood Foundation Child Health Research Award, the Glass Family Charitable Foundation, and NIH grants 1R56AI114525 and 1R21AI119885.

References

1. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826. doi:10.1126/science.1232033
2. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6):1262–1278. doi:10.1016/j.cell.2014.05.010
3. Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482(7385):331–338. doi:10.1038/nature10886
4. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819):1709–1712. doi:10.1126/science.1138140
5. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32(4):347–355. doi:10.1038/nbt.2842
6. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. doi:10.1126/science.1231143
7. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation

and editing. Nat Methods 10(11):1116–1121. doi:10.1038/nmeth.2681

8. Price AA, Sampson TR, Ratner HK, Grakoui A, Weiss DS (2015) Cas9-mediated targeting of viral RNA in eukaryotic cells. Proc Natl Acad Sci U S A 112(19):6164–6169. doi:10.1073/pnas.1422340112
9. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature 520(7546):186–191. doi:10.1038/nature14299
10. Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513(7519):569–573. doi:10.1038/nature13579
11. Zhang Y, Ge X, Yang F, Zhang L, Zheng J, Tan X, Jin ZB, Qu J, Gu F (2014) Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci Rep 4:5405. doi:10.1038/srep05405
12. Kiani S, Beal J, Ebrahimkhani MR, Huh J, Hall RN, Xie Z, Li Y, Weiss R (2014) CRISPR transcriptional repression devices and layered circuits in mammalian cells. Nat Methods 11(7):723–726. doi:10.1038/nmeth.2969
13. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41(15):7429–7437. doi:10.1093/nar/gkt520
14. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154(2):442–451. doi:10.1016/j.cell.2013.06.044
15. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152(5):1173–1183. doi:10.1016/j.cell.2013.02.022
16. Kearns NA, Genga RM, Enuameh MS, Garber M, Wolfe SA, Maehr R (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development 141(1):219–223. doi:10.1242/dev.103341
17. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods 10(10):973–976. doi:10.1038/nmeth.2600
18. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc 8(11):2180–2196. doi:10.1038/nprot.2013.132
19. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat Methods 12(5):401–403. doi:10.1038/nmeth.3325
20. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33(5):510–517. doi:10.1038/nbt.3199
21. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS (2014) Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159(3):647–661. doi:10.1016/j.cell.2014.09.029
22. Echeverri CJ, Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. Nat Rev Genet 7(5):373–384. doi:10.1038/nrg1836
23. Li X, Krawetz R, Liu S, Meng G, Rancourt DE (2009) ROCK inhibitor improves survival of cryopreserved serum/feeder-free single human embryonic stem cells. Hum Reprod 24(3):580–589. doi:10.1093/humrep/den404
24. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31(9):822–826. doi:10.1038/nbt.2623
25. Chou YC, Lai MM, Wu YC, Hsu NC, Jeng KS, Su WC (2015) Variations in genome-wide RNAi screens: lessons from influenza research. J Clin Bioinforma 5:2. doi:10.1186/s13336-015-0017-5
26. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31(9):827–832. doi:10.1038/nbt.2647

Chapter 17

In Vitro Assay to Study Histone Ubiquitination During Transcriptional Regulation

Jogender Tushir-Singh and Sanchita Bhatnagar

Abstract

In mammals, gene expression is largely controlled at the transcriptional level. In response to environmental or intrinsic signaling, gene expression is often fine-tuned by epigenetic modifications, including DNA methylation and histone modifications. One such histone modification is ubiquitination that predominately occurs in mono-ubiquitinated forms on histone H2A and H2B. We recently identified and characterized a novel E3 ligase called TRIM37 that ubiquitinates H2A. This study highlights the consequence of aberrant histone ubiquitination at the promoters of tumor suppressor genes in breast cancer. Regulatory mechanism by which TRIM37 and other auxiliary proteins are involved in the initiation and progression of breast cancer is of utmost importance toward generating effective therapeutics. Here, we describe a detailed step-by-step process of carrying out in vitro ubiquitination assay using purified histone proteins or reconstituted nucleosomes and affinity-purified recombinant E3 ligase like TRIM37. These experimental procedures are largely based on our studies in mammalian cells and will be a useful tool to identify substrate for E3 ubiquitin ligase as well as characterizing new E3 ligases.

Key words Ubiquitination, In vitro assay, Histones, H2A, Nucleosomes, Recombinant protein, E3 ubiquitin ligase

1 Introduction

Chromatin is a macromolecular complex for genomic DNA, and nucleosome represents a fundamental building block for chromatin. Each nucleosome represents "bead on a string" model where two turns of DNA (~146 bp) are wrapped around a histone octamer. Each core histone octamer is composed of four histone proteins (H2A, H2B, H3, and H4) that are linked to an adjacent nucleosome by a linker histone protein H1 [1]. Often dynamic changes in chromatin conformations and transcriptional regulations occur by two major epigenetic mechanisms: (1) chromatin-remodeling complexes and (2) DNA and histone tail modifications.

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_17,

Chromatin-remodeling complexes tailor the dynamic interactions in the chromatin by altering nucleosomal compositions [2]. CpG methylation is one of the major covalent DNA modifications [3], whereas histones can undergo covalent post-translational modifications (PTMs) that include acetylation, phosphorylation, methylation, ubiquitination, and poly(ADP-ribosylation) [4, 5]. Different modifications affect transcription differentially, for example, trimethylation of histone H3 at Lys 4 (H3K4me3) is abundant in actively transcribed region, whereas trimethylation of H3 at Lys 9 (H3K9me3) and trimethylation of histone H3 at Lys 27 (H3K27me3) represent transcriptionally silenced region [6]. These modifications are mediated by methyltransferases, and several of them play crucial role in many biological processes, including cell-cycle regulation, DNA damage, and stress response. Importantly, there is increased interest in the methyltransferases due to their link to cancer and aging [7]. Acetylation of histones is also associated with actively transcribed chromatin, for example, histone H3 Lys4, and is mediated by histone acetyltransferase (HAT) [8].

Another important covalent histone modification is the mono-ubiquitination that occurs on histone H2A and H2B. Mono-ubiquitination of H2B affects transcription both positively and negatively [9], while mono-ubiquitination of H2A at Lys 119 is associated with the transcriptional repression [10]. A mono-ubiquitination event is catalyzed by three different enzymes, in a step-wise process – initiated by the ATP-dependent ubiquitin-activating enzyme (E1), followed by ubiquitin transferase (E2), and terminal ubiquitin ligase (E3) (Fig. 1).

One of the major mammalian E3 ubiquitin ligases is RNF2 that functions in association with a multi-subunit polycomb

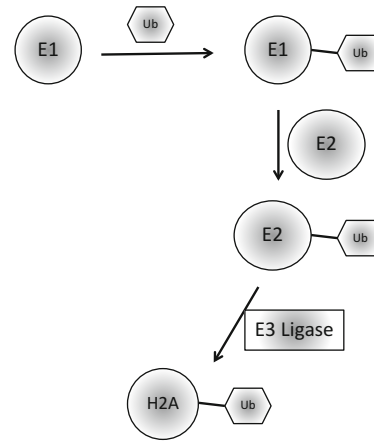

Fig. 1 Schematic illustrating the mono-ubiquitination pathway. E1 enzyme binds ubiquitin in ATP-dependent manner and transfers to E2 ubiquitin-conjugating enzyme. E3 enzyme facilitates the transfer of ubiquitin moiety from E2 to the specific substrate

complex, PRC1 [11]. Recently, we discovered a new H2A ubiquitin E3 ligase called TRIM37 that is overexpressed in breast cancer carrying amplification of 17q23 region [12]. We showed that TRIM37 functions as an oncogene by promoting ubiquitination at the promoters of several tumor suppressor genes identified by genome-wide (ChIP)-chip analysis and, consequently, renders a cell tumorigenic. In the following section, we describe a step-by-step detailed protocol for in vitro ubiquitination assays using H2A as a substrate. We believe the protocol outlined here will be instrumental in identifying and characterizing new ubiquitin E3 ligases.

2 Materials

All buffer solutions are prepared using autoclaved distilled water and stored at room temperature unless stated otherwise. All reagents used in the protocol are ACS grade.

2.1 Recombinant TRIM37 Production

Cells: COS-1 cells (ATCC).

Cell culturing conditions: Cells are maintained in ATCC-formulated Dulbecco's Modified Eagle's Medium (ATCC) supplemented with 10% fetal bovine serum at 37 °C, 5% CO_2.

Constructs: TRIM37 cDNA (7) is cloned in a derivative of pEF6/V-5HisB resulting in addition of protein C-epitope (MAEDQVDPRLIDGKEFT) at the N-terminus. A TRIM37 derivative carrying mutation was generated by PCR with overlapping primers [12].

Reagents: Effectene transfection reagent (Qiagen), anti-protein C affinity matrix, high-fidelity *Pfu* polymerase.

2.2 Expression of Recombinant Histone H2A

Constructs: A PET-based plasmid expressing H2A (a gift from Dr. Craig Peterson, University of Massachusetts Medical School) was used as a template in a PCR-based site-directed mutagenesis with primer extension reaction using primers ATGTCTGGTCGTGGCAA ACA, ATGCGCGTCCTCCTGTTGTC, (P)-CATCCCTCGTCACC TCCAG, and TGGGTGGCTCTAAAAAGAGCC (7).

Reagents: IPTG, ampicillin, chloramphenicol, PMSF, DTT, benzamidine, DMSO, HA peptide, β-mercaptoethanol (BME), dialysis membrane 3500 MWCO.

Buffers: 2×YT (tryptone, yeast extract NaCl).

Histone Wash T: 50 mM Tris–HCl, pH 7.4, 1 mM EDTA, 100 mM NaCl, 1 mM benzamidine and 1 mM DTT.

Histone Wash TW: 50 mM Tris–HCl, pH 7.4, 1 mM EDTA, 100 mM NaCl, 1% (v/v) Triton X-100; add 0.5 mM PMSF, 1 mM Benzamidine, and 1 mM DTT immediately before use.

Unfolding buffer: 20 mM Tris–HCl, pH 7.4, 10 mM DTT, 7 M guanidine HCl. Pass through 0.4 μm filters before use.

Urea dialysis buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 7 M urea. Add 0.5 mM PMSF and 5 mM BME immediately before use.

Urea low buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 7 M urea. Make immediately before use and pass through 0.45 μm filters. Add 0.5 mM PMSF and 1 mM DTT immediately before use.

Urea high buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 1 M NaCl, 7 M urea. Add 0.5 mM PMSF and 1 mM DTT immediately before use.

2.3 In Vitro Ubiquitination Assay

Reagents: H2A or H2B (5×), UBE1, UBCH5B, and HA-ubiquitin aldehyde and ATP.

Buffers: Prepare the reaction buffer immediately before use. 50 mM Tris–HCl, pH 7.9, 5 mM $MgCl_2$, 2 mM sodium fluoride, 50 mM dithiothreitol (DTT), 0.5 mM PMSF, 5 μg/mL aprotinin, 5 μg/mL leupeptin, and 5 μg/mL pepstatin.

2.4 Equipment

SDS-PAGE apparatus, electrophoresis apparatus, transfer unit, power supply, nitrocellulose membrane (0.2 μm), sonicator.

3 Methods

3.1 Recombinant TRIM37 Expression

1. TRIM37 open reading frame was PCR amplified from a cDNA clone using high-fidelity *Pfu* polymerase and cloned into a derivative of pEF6/V-5HisB.
2. COS-1 cells were transfected using Effectene transfection reagent according to manufacturer's protocol. Optimal protein expression and cell viability require optimizing transfection conditions.
3. Transiently expressing cells were lysed after 48 h, and cell extract were prepared.
4. Anti-protein C-tagged affinity matrix was used to enrich the expressed protein.

3.2 Expression of Recombinant Histone H2A

Adapted from previously published protocol [13]. This protocol is optimized for 2 L of bacterial cultures, and the volume should be adjusted according to the final yield expected (*see* **Note 1**).

1. Streak out desired BL-21 (DE3) pLysS strain (CP812xH2A-HAwt (a gift from Dr. Craig Peterson, University of Massachusetts Medical School) or CP812xH2A KKRR-HA (7)) on freshly prepared LB plates containing ampicillin and chloramphenicol.
2. Inoculate two 5 mL starter 2×YT cultures (containing ampicillin and chloramphenicol) with a colony from a plate and incubate at 37 °C on shaking platform (200–240 × *g*). 8 h later,

subculture 50 mL of 2×YT media with 1 mL of the inoculum and incubate overnight at 37 °C without rotation.

3. Use 10 mL of inoculum from overnight-grown culture to inoculate four 500 mL of 2×YT (containing chloramphenicol and ampicillin) media in 1-L culture flasks. Grow into log phase at 37 °C on rotation platform (200–240 × *g*) until absorbance reaches (0.6).
4. Induce cultures for 3 h by adding 0.2 mM IPTG (*see* **Note 2**).
5. After 3 h, harvest cells by centrifugation at 3000 × *g* for 30 min at room temperature.

3.3 Purification of Recombinant Histones

The purification of histone H2A or H2A-KKRR (H2A derivative with mutation at Lys119) involves three steps:

1. Inclusion body preparation in denaturing condition
2. Refolding of protein
3. Affinity purification

3.3.1 Inclusion Body Preparation

1. Resuspend pellet, obtained in Subheading 3.2, in 16.7 mL Histone Wash T and flash freeze (*see* **Note 3**).
2. Thaw pellet in 30 °C water bath and by mixing occasionally. Adjust volume to 25 mL for previously resuspended pellet with fresh Histone Wash T buffer. Handle the pellet on ice this step onward.
3. Sonicate cells at frequency setting of 5 for 15 s on ice. Continue until the solution is no longer viscous (*see* **Note 4**).
4. Centrifuge sonicated cells in JA-17 rotor at 14,000 × *g* for 20 min at 4 °C.
5. Reconstitute the pellet using Histone Wash TW buffer. Adjust the volume to 25 mL per pellet from a 1-L culture (*see* **Note 4**).
6. Discard the supernatant.
7. The resultant pellet represents the "inclusion body." The inclusion body when thawed should be whitish and the size of the pellet reflects the efficiency of induced histone. A little black or gray discoloration is to be expected.

3.3.2 Refolding the Histone H2A Protein

1. Thaw the inclusion body at room temperature.
2. Resuspend the pellet in 350 μl of DMSO and incubate at room temperature for 30 min. Pellet can be resuspended using a small magnetic stirrer.
3. Combine both 1 L pellets together into one centrifuge tube using a spatula.
4. Add 26.6 mL fresh unfolding buffer to resuspend the pellet. Use spatula or a stir bar to break the pellet as much as possible.

5. Harvest cells at 14,000 × *g* for 20 min at rom temperature and collect the supernatant (*see* **Note 5**).
6. To maximize the yield, re-extract the pellet with 9.4 mL unfolding buffer after resuspending pellet for 30 min at room temperature. Centrifuge at 14,000 × *g* to remove debris and collect the supernatant.
7. Pool the supernatants obtained in **steps 5** and **6** and dialyze against 2 L of 7 M urea dialysis buffer at 4 °C for 3 h.
8. Repeat the dialysis in **step 7** at 4 °C overnight.

3.3.3 Affinity Purification

1. Next day, carefully collect the dialyzed solution and diluted with Tris–HCl, pH 7.0 and incubate with anti-HA matrix at 4 °C for 3–4 h.
2. Wash with at least fourfold volumes of buffer (4× buffer: 1× beads).
3. Use HA peptide to elute the matrix bound protein.
4. Get the concentration of the histone protein post-dialysis by reading absorption at 276 nm (*see* **Note 6**).
5. Aliquot into volumes that are convenient for reconstitution (100 nmol). Freshly prepared histones can be used immediately or can be stored at −20 °C for short term. For long-term use, aliquots should be snap-frozen, lyophilized, and stored at −80 °C (*see* **Note 7**).

3.4 In Vitro Histone Ubiquitination Assay with Purified Histones

1. Mix 1× histone proteins (H2A/H2B) or 5 μg of H2A-HA or 5 μg of H2A-KKRR-HA; 10 nM ubiquitin-activating enzyme (UBE1); 10 nM UBCH5A, UBCH5B, or UBCH5C; and 10 nM HA-ubiquitin aldehyde in 10 μl of the reaction buffer (*see* **Note 8**).
2. Add 5 μg of purified TRIM37 or TRIM37 derivative with point mutation C18R.
3. Add 2 mM of ATP and adjust the total volume to 10 μl with the reaction buffer.
4. Incubate the reaction at 32 °C for 90 min.
5. Stop the reaction by adding SDS sample loading buffer.
6. Proteins were resolved on 15% SDS-PAGE and transferred to a nitrocellulose membrane (*see* **Note 9**).
7. Transferred proteins are subjected to immunoblotting with H2AubK119, H2A, H2bub, and H2B (Fig. 2) [14].
8. Negative control assays: Omit the substrate (ubiquitin) or putative or known E3 ligase or E1 or E2 in the assay and run it parallel in the experiment. Another control used in the study is histone H2A derivative with mutation in K119 site, and loss of ubiquitination site should eliminate the H2A ubiquitination signal (Fig. 2) [14].

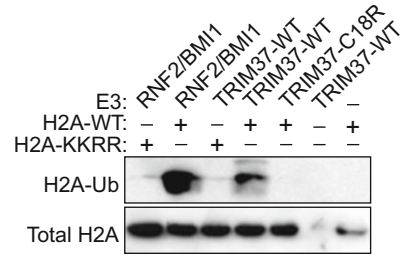

Fig. 2 In vitro ubiquitination assay with recombinant H2A. In vitro ubiquitination assay done in the presence of E1, E2, and TRIM37 as ubiquitin E3 ligase. Immunoblots probed with either ubiquityl-H2A or H2A antibody. Recombinant H2A or mutant H2A (KKRR) was used as a substrate in the assay. Recombinant RNF2 and BMI1 were used as a positive control as described in the text

9. Positive control assays: RNF2/BMI1 are known H2A E3 ligase and used as positive control for H2A ubiquitination (Fig. 2). BRCA1 is a known H2B E3 ligase and used as a positive control for studying H2B ubiquitination [14].

3.5 In Vitro Ubiquitination Assay with Reconstituted Nucleosomes

1. Mix 10 nM ubiquitin-activating enzyme (UBE1); 10 nM UBCH5A, UBCH5B, or UBCH5C; and 10 nM HA-ubiquitin aldehyde.
2. 5 μg of *Xenopus* oocyte-derived nucleosomes were added to the reaction mixture (*see* **Note 10**).
3. Add 5 μg of purified TRIM37 or TRIM37 derivative with point mutation C/R.
4. Add 2 mM of ATP and adjust the total volume to 10 μl with the reaction buffer.
5. Reactions were incubated at 32 °C for 90 min.
6. Stop the reaction by adding SDS sample loading buffer.
7. Proteins were resolved on 15% SDS-PAGE and transferred to a nitrocellulose membrane.
8. Transferred proteins are subjected to immunoblotting with H2AubK119, H2A, H2bub, H2B, H3, and H4 (Fig. 3) [14].
9. Negative control assays: Omit the substrate (ubiquitin) or putative or known E3 ligase or E1 or E2 in the assay and run it parallel in the experiment. This should eliminate the H2A ubiquitination signal (Fig. 3) [14].
10. Positive control assays: RNF2/BMI1 are known H2A E3 ligase and used as positive control for H2A ubiquitination (Fig. 3). BRCA1 is a known H2B E3 ligase and used as a positive control for studying H2B ubiquitination.

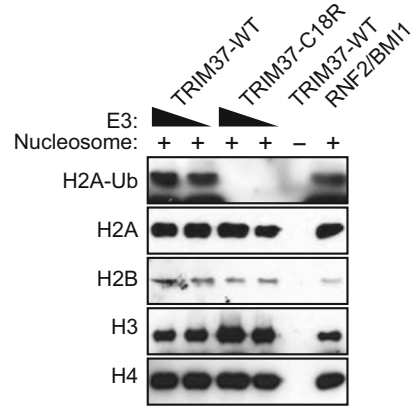

Fig. 3 In vitro ubiquitination assay with reconstituted nucleosomes as substrates. In vitro ubiquitination assay done in the presence of E1, E2, and TRIM37 as ubiquitin E3 ligase or TRIM37 derivative carrying mutation (C18R). Immunoblots probed with either ubiquityl-H2A, H2A, H2B, H3, or H4 antibody. Recombinant RNF2 and BMI1 were used as a positive control as described in the text

4 Notes

1. In the purification protocol, we did not use the tandem ion-exchange columns for purification of histones, as described previously [14].
2. Remember to take pre- and post-induction time points to evaluate the expression of histones. Harvest cells from 1 mL of culture by centrifuging at 14,000 × *g* for 5 min at 4 °C. Resuspend cell pellet in 200 μL of 2×SDS-PAGE sample buffer and boil samples for 10 min before analyzing on 15% SDS-PAGE. A good induction will yield a very strong band.
3. The color of the pellet after induction indicates the quality of induction; good induction cultures yield much whiter bacterial pellets. If larger amounts of protein are required, the culture volume should be increased accordingly.
4. If solution contains yellowish chunks after sonication or is still viscous, cell pellet can be sonicated for longer time. If supernatant is viscous after centrifugation, repeat the sonication (**step 3**) and resuspend pellet once more in Histone Wash TW.
5. The volume of unfolding buffer can be adjusted according to the yield of protein expected.
6. Use the last change of dialysis water as a blank when measuring the absorbance of the histones. To accurately determine histone concentration, it is important to use at least three different dilutions of purified histone. The protocol outlined here can be used to express and purify other HA-tagged histone proteins with the molecular weight for respective histone protein listed in the table below:

Histone	E (276 nm, per cm per M)	Molecular weight (Da)
H2A	4050	13,960
H2B	6070	13,774
H3	4040	15,273
H4	5400	11,236

7. Store the affinity-purified proteins at −80 °C and avoid repeated freeze thaw.
8. We prefer to use freshly prepared for in vitro ubiquitination assay.
9. The nitrocellulose membrane works better compared to PVDF for histone transfer. We used 0.2 μm membrane for Western blotting.
10. We prefer to use freshly reconstituted nucleosomes as a substrate for the in vitro ubiquitination assay. Nucleosomes can be stored at 4 °C for 2 weeks before using for ubiquitination assay, and precipitation is seen with long-term storage that can interfere with the assay.

Acknowledgments

We thank Dr. Craig Peterson (UMASS Medical School) for providing reconstituted nucleosomes and plasmid for histone H2A expression. We would also like to thank Dr. Michael R. Green (UMASS Medical School) who provided reagents and resources for work outlined here.

References

1. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389(6648):251–260. doi:10.1038/38444
2. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78:273–304. doi:10.1146/annurev.biochem.77.062706.153223
3. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14(3):204–220. doi:10.1038/nrg3354, nrg3354 [pii]
4. Shilatifard A (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem 75:243–269. doi:10.1146/annurev.biochem.75.103004.142422
5. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21(3):381–395. doi:10.1038/cr.2011.22, cr201122 [pii]
6. Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13(5):343–357. doi:10.1038/nrg3173, nrg3173 [pii]
7. Albert M, Helin K (2010) Histone methyltransferases in cancer. Semin Cell Dev Biol 21(2):209–220. doi:10.1016/j.semcdb.2009.10.007, S1084-9521(09)00199-2 [pii]
8. Verdone L, Caserta M, Di Mauro E (2005) Role of histone acetylation in the control of gene expression. Biochem Cell Biol 83(3):344–353. doi:10.1139/o05-041, o05-041 [pii]
9. Espinosa JM (2008) Histone H2B ubiquitination: the cancer connection. Genes

Dev 22(20):2743–2749. doi:10.1101/gad.1732108, 22/20/2743 [pii]

10. Zhou W, Wang X, Rosenfeld MG (2009) Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. Int J Biochem Cell Biol 41(1):12–15. doi:10.1016/j.biocel.2008.09.016, S1357-2725(08)00393-2 [pii]
11. Zhang Y (2003) Transcriptional regulation by histone ubiquitination and deubiquitination. Genes Dev 17(22):2733–2740. doi:10.1101/gad.1156403, 17/22/2733 [pii]
12. Bhatnagar S, Green MR (2015) TRIMming down tumor suppressors in breast cancer. Cell Cycle 14(9):1345–1346. doi:10.1080/15384101.2015.1024584
13. Luger K, Rechsteiner TJ, Richmond TJ (1999) Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol 119:1–16. doi:10.1385/1-59259-681-9:1, 1-59259-681-9-1 [pii]
14. Bhatnagar S, Gazin C, Chamberlain L, Ou J, Zhu X, Tushir JS, Virbasius CM, Lin L, Zhu LJ, Wajapeyee N, Green MR (2014) TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. Nature 516(7529):116–120. doi:10.1038/nature13955, nature13955 [pii]

Chapter 18

Determination of Alternate Splicing Events Using Transcriptome Arrays

Rebecca S. DeVaux, Marcy Kuentzel, Jason Herschkowitz, and Sridar V. Chittur

Abstract

Understanding differential isoform expression is critical to mechanistically illuminate the biology underlying both normal development and disease states. High-throughput expression profiling analysis of splice variants has thus far been limited by sample requirements and an appropriate platform for quantitation and analysis. Here we describe Affymetrix GeneChip Human Transcriptome Array 2.0, which is employed for comprehensive examination of all known transcript isoforms.

Key words Alternative splicing, Gene regulation, Expression profiling, Microarray, Exon splicing

1 Introduction

Sequencing of the human genome is one of the most significant scientific accomplishments of our time. When the human genome was released in 2001, it was surprising that all of the diversity encompassed in human biology was captured in only ~21,000 genes [1, 2]. This was especially startling when the genomes of far less complex organisms, such as the nematode *Caenorhabditis elegans* and common mouse, were found to have ~19,000 [3] and ~22,000 [4] genes, respectively. It quickly became apparent that the number of encoded genes alone could not explain the complexity of human biology.

Following the human genome project, the encyclopedia of DNA elements (ENCODE) project was initiated to dissect the genome and identify functional elements [5]. That ambitious enterprise revealed that only approximately 2% of the human genome encodes for protein sequences. The exons encoding for the translated messenger RNA (mRNA) were found to be interspersed between noncoding introns. To mediate the split nature of eukaryotic genes, a spliceosome complex functions to remove

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_18,

introns and appropriately join coding exons and form mRNA. As opposed to the initial hypothesis that one gene results in one protein, now it is well accepted that alternative splicing of a locus can result numerous isoforms. In fact, alternative splicing is conservatively estimated to occur in ~94% of human genes. Misregulated alternative splicing is estimated to contribute to as many as 50% of human genetic diseases and further promote tumorigenesis [6]. Thus, to decipher the genetic alterations underlying disease states, we must understand not only which genes are altered, but we must pinpoint exactly which isoforms are affected.

In the study described herein, we use the Affymetrix GeneChip® Human Transcriptome Array 2.0 (HTA 2.0) [9] to comprehensively examine the isoforms of both coding and noncoding transcripts altered due to expression of a putative driver of breast cancer progression.

2 Materials and Reagents

2.1 Equipment

1. Agilent Bioanalyzer 2100 system.
2. NanoDrop ND-1000 UV-Vis spectrophotometer.
3. Affymetrix GeneChip® System with FS450 and GCS3000 7G Scanner.

2.2 Materials for Cell Culture

1. MCF10DCIS.com cells were maintained in MCF10A media—(DMEM/F12, 5% horse serum, 20 ng/mL EGF, 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 μg/mL insulin, 1× antibiotic-antimycotic).
2. MCF10DCIS.com cells were reverse transfected using RNAiMAX (Life Technologies) and 50 nM of either Silencer® Select Negative control No. 1 siRNA (Life Technologies) or siRNA targeting BHLHE40-AS1.

2.3 Materials for RNA Isolation

1. All tips, tubes, and reagent bottles must be DNAse- and RNase-free (*see* **Note 1**).
2. Omega Bio-tek E.Z.N.A.® Total RNA Kit I. (Omega Bio).
3. We recommend the use of nuclease-free water to prepare all buffers and solutions.
4. RNaseZap.

2.4 Materials for RNA QC and Microarray Experiment

1. RNA 6000 Nanokit (Agilent).
2. GeneChip WT PLUS Reagent kit (Affymetrix).
   This catalog number includes all kits required for this protocol including cDNA synthesis, amplification, labeling, cleanup, and hybridization.

3. GeneChip® Human Transcriptome 2.0 ST arrays (Affymetrix).
4. Magnetic stand-96.

3 Methods

3.1 Cell Culture and Harvesting of Cells for RNA Isolation

1. Indicated cell line was cultured in MCF10A media and maintained in 10-cm^2 dishes in a 37 °C incubator under 5% CO_2.
2. Cells were reverse transfected with siRNA and seeded into 6-well plates such that 24 h post transfection, cells were approximately 50% confluent. Cells were harvested 72 h post transfection when cells were approximately 85% confluent (*see* **Note 2**).
3. Wash the cells with PBS to remove any residual media prior to harvesting.
4. Lyse cells according to selected RNA isolation method or kit (*see* Subheading 3.2).
5. Alternatively, add TRIzol Reagent directly to the cells. One mL per 10-cm^2 dish, 200 µL per 6 wells. *Do not* trypsinize the cells prior to treatment with tri-reagent or TRIzol. Move the TRIzol around the dish and gently tap to slough off all attached cells. Pipette into a clean tube and store at –20 °C till further use.

3.2 RNA Isolation

The specific RNA isolation method that you choose will depend on your downstream application. Generally most methods are acceptable for microarray, RT-PCR, or Northern blotting. The Qiagen RNeasy and Omega Bio-tek spin-column isolation kits offer the advantage of performing an optional DNase I digestion while purifying the RNA, so further processing is avoided. However, detection of RNA molecules of 200 bp or smaller will be limited if using the Omega cleanup procedure and hence not advised if you intend to use the RNA for miRNA analysis (*see* **Note 3**). Alternative spin-column kits specifically indicated to capture miRNA can be purchased. While using arrays such as the transcriptome ST or Gene ST arrays, ensure that the RNA is DNase treated since DNA contaminants will be amplified and labeled in the array protocol.

3.3 Assessment of RNA Quality

1. Using a NanoDrop® spectrophotometer, measure the optical absorbance characteristics of the sample. The $A_{260/A280}$ as well as the $A_{260/A230}$ ratio will ideally be close to 2.0, signifying the purification of nucleic acids away from protein and other organics, respectively. If either ratio is lower than 1.6, expect problems with downstream applications of the RNA (*see* **Note 4**).
2. Performance of a NanoChip assay using Agilent's Bioanalyzer allows for measurement of the molecular weight profile of the

isolated RNA. In this way, you may evaluate the 28S/18S ratio measurements. A total RNA ratio between 1.8 and 2.0 is desirable; however, ratios 1.6–1.8 may be acceptable. A RNA integrity number (RIN) score should be between 7 and 10 if the samples are to be used in a microarray or QPCR experiment downstream (*see* **Note 5**).

3.4 Expression Analysis of mRNA from Cells

Affymetrix recommends amplification and labeling of the RNA for hybridization to transcriptome arrays starting with at least 50 ng of total RNA. We will demonstrate use of the 100 ng protocol in this example (*see* Fig. 1) [8] (*see* **Note 6**). We have had good results with this protocol and also with the Affymetrix WT Pico protocol, which enables starting with much lower amounts of RNA (100 pg to 1 ng) as seen with LCM or flow-sorted samples. Please remember that since data generated by each of these protocols are not directly cross comparable, process all samples of a given study using the same protocol.

3.5 Synthesis of Labeled cDNA and Microarray Hybridization

3.5.1 Preparation of cRNA

1. Make serial dilutions of the GeneChip PolyA controls (1:20; 1:50, 1:50, and finally 1:10) using the PolyA control dilution buffer supplied with the kit. The final concentration of the PolyA controls is 1:500,000 of the original stock.
2. Add 2 μl of the diluted poly A controls to the 100 ng total RNA in a RNAse free tube. The total volume of the mixture should not exceed 5 μl.
3. Prepare the first strand master mix composed of 4 μl first-strand buffer and 1 μl of the first-strand enzyme per reaction. Mix thoroughly by vortexing, centrifuge briefly and place on ice. To this tube, add 5 μl of the total RNA from **step 2**. Mix thoroughly by vortexing, centrifuge briefly and then incubate at 1 hr at 25 °C, then for 1 hr at 42 °C, then for at least 2 min at 4 °C. Proceed immediately to **step 4**. (*see* **Note 7**).
4. Prepare the second-strand master composed of 18 μl second-strand buffer and 2 μl second-strand enzyme per reaction. Mix thoroughly by vortexing, centrifuge briefly and place on ice. Transfer 20 μl of this second-strand master mix to the first-strand cDNA from **step 3**. Mix thoroughly by vortexing, centrifuge briefly and then incubate for 1 hr at 16 °C, then for 10 min at 65 °C, then for at least 2 min at 4 °C (*see* **Note 8**).
5. Prepare the IVT master composed of 24 μl IVT buffer and 6 μl IVT enzyme per reaction. Mix thoroughly by vortexing, centrifuge briefly and immediately transfer 30 μl of this IVT master mix to the second-strand cDNA from step 4. Mix thoroughly by vortexing, centrifuge briefly and then incubate for 16 hr at 40 °C, then at 4 °C (*see* **Note 9**).

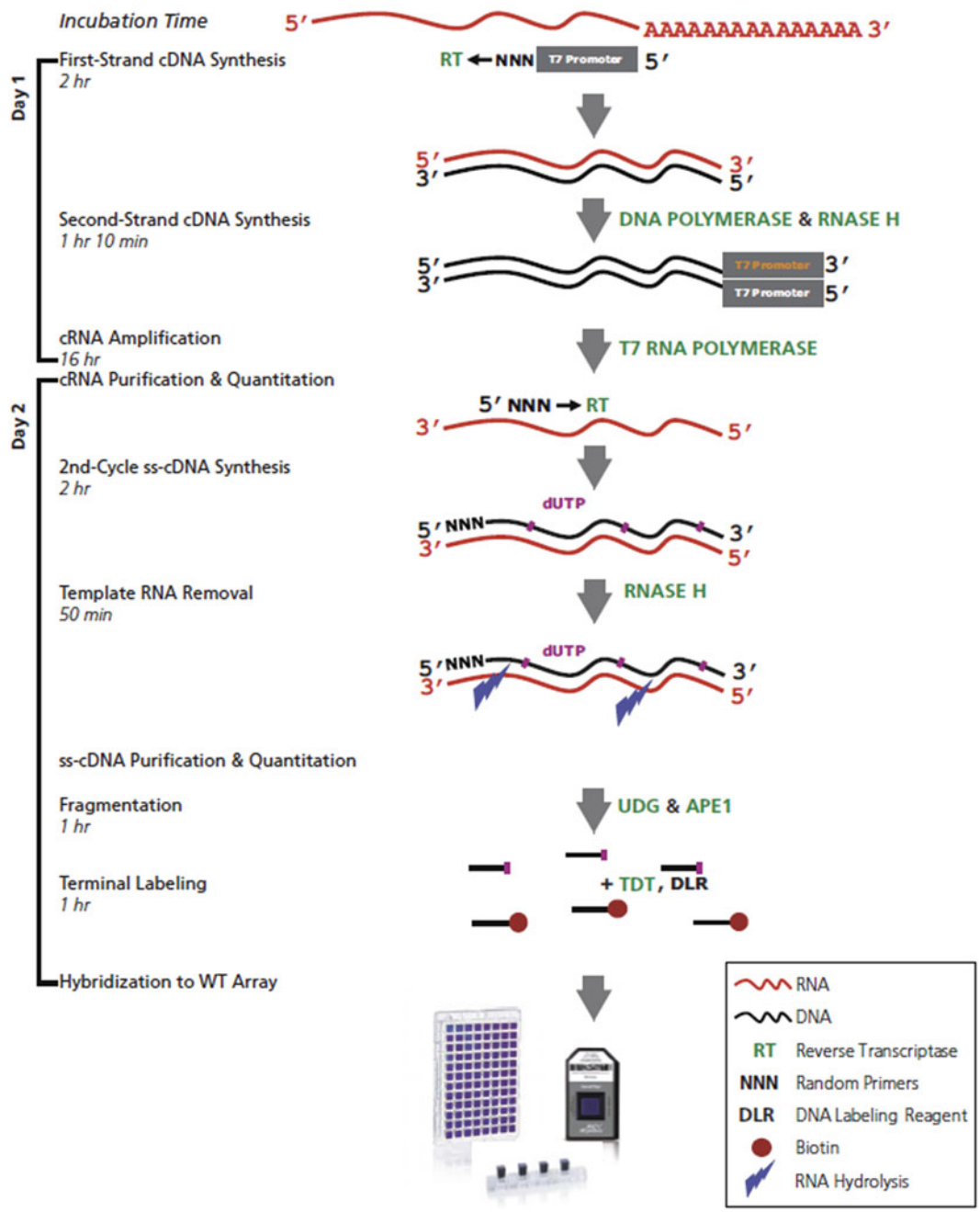

Fig. 1 WT PLUS amplification and labeling workflow [8]

6. After the incubation, centrifuge briefly to collect the cRNA at the bottom of the tube or well. Place the reaction on ice, then proceed to purify cRNA, or immediately freeze the samples at −20 °C for storage.
7. Mix the purification beads by vortexing to resuspend the magnetic particles and then add 100 μl to each (60 μl) cRNA

sample from **step 6**, mix by pipetting and transfer to a well of a U-bottom plate (*see* **Note 10**).

8. Mix well by pipetting up and down ten times and then incubate for 10 min.
9. Move the plate to a magnetic stand to capture the purification beads. When capture is complete (after ~5 min), the mixture is transparent, and the purification beads form pellets against the magnets in the magnetic stand. The exact capture time depends on the magnetic stand that you use and the amount of cRNA generated by in vitro transcription. Carefully aspirate and discard the supernatant without disturbing the purification beads.
10. While on the magnetic stand, add 200 µL of 80 % ethanol wash solution to each well and incubate for 30 s. Slowly aspirate and discard the 80 % ethanol wash solution without disturbing the purification beads. Repeat twice for a total of three washes with 200 µL of 80 % ethanol wash solution. Completely remove the final wash solution.
11. Air-dry on the magnetic stand for 5 min until no liquid is visible, yet the pellet appears shiny (*see* **Note 11**).
12. Remove the plate from the magnetic stand. Add to each sample 27 µL of the preheated (65 °C) nuclease-free water and incubate for 1 min. Mix well by pipetting up and down ten times, and place the plate on the magnetic stand for ~5 min to capture the purification beads. Transfer the supernatant, which contains the eluted cRNA, to a nuclease-free tube.
13. Place the purified cRNA samples on ice, then proceed to assess cRNA yield and size distribution, or immediately freeze the samples at −20 °C for storage.

3.5.2 Preparation of Second-Cycle Single-Stranded cDNA

1. On ice, combine 15 µg cRNA (in a volume of 24 µL) and 4 µL of second-cycle primers. Mix thoroughly by vortexing, centrifuge briefly, and then incubate for 5 min at 70 °C, 5 min at 25 °C, and then 2 min at 4 °C.
2. Prepare the second-cycle ss-cDNA master mix composed of 8 µL second-cycle ss-cDNA buffer, and 4 µL of the second-cycle ss-cDNA enzyme per reaction. Mix thoroughly by vortexing, centrifuge briefly, and place on ice. On ice, transfer 12 µL of the second-cycle ss-cDNA master mix to each (28 µL) cRNA/second-cycle primer sample for a final reaction volume of 40 µL. Mix thoroughly by gently vortexing the tube, centrifuge briefly, and incubate for 10 min at 25 °C, then 90 min at 42 °C, then 10 min at 70 °C, then for at least 2 min at 4 °C.
3. Add 4 µL of the RNase H to each (40 µL) second-cycle ss-cDNA sample for a final reaction volume of 44 µL. Mix thoroughly. Centrifuge briefly and then incubate for 45 min at 37 °C, then for 5 min at 95 °C, then for at least 2 min at 4 °C.

4. After incubation, centrifuge briefly and place the tubes on ice. Add 11 μL of the nuclease-free water to each (44 μL) hydrolyzed second-cycle ss-cDNA sample for a final reaction volume of 55 μL. Mix well, centrifuge, and chill on ice before purifying the second-cycle ss-cDNA.
5. Mix the purification beads by vortexing to resuspend the magnetic particles and then add 100 μL to each (55 μL) ss-cDNA sample from **step 4**; mix by pipetting and transfer to a well of a U-bottom plate. Add 150 μL of 100% ethanol to each (155 μL) ss-cDNA/bead sample (*see* **Note 10**).
6. Mix well by pipetting up and down ten times and incubate for 20 min.
7. Move the plate to a magnetic stand to capture the purification beads. When capture is complete (after ~5 min), the mixture is transparent, and the purification beads form pellets against the magnets in the magnetic stand. The exact capture time depends on the magnetic stand that you use and the amount of ss-cDNA generated by in vitro transcription. Carefully aspirate and discard the supernatant without disturbing the purification beads.
8. While on the magnetic stand, add 200 μL of 80% ethanol wash solution to each well and incubate for 30 s. Slowly aspirate and discard the 80% ethanol wash solution without disturbing the purification beads. Repeat twice for a total of three washes with 200 μL of 80% ethanol wash solution. Completely remove the final wash solution.
9. Air-dry on the magnetic stand for 5 min until no liquid is visible, yet the pellet appears shiny (*see* **Note 11**).
10. Remove the plate from the magnetic stand. Add to each sample 30 μL of the preheated (65 °C) nuclease-free water and incubate for 1 min. Mix well by pipetting up and down ten times and place the plate on the magnetic stand for ~5 min to capture the purification beads. Transfer the supernatant, which contains the eluted ss-cDNA, to a nuclease-free tube.
11. Place the purified ss-cDNA samples on ice, then proceed to assess ss-cDNA yield and size distribution, or immediately freeze the samples at −20 °C for storage.

3.5.3 Preparation of Fragmented and Labeled Single-Stranded cDNA

1. On ice, prepare 5.5 μg ss-cDNA in a volume of 31.2 μL.
2. Prepare the fragmentation master mix composed of 10 μL nuclease-free water, 4.8 μL 10× cDNA fragmentation buffer, 1 μL uracil-DNA glycosylase, and 1 μL apurinic/apyrimidinic endonuclease 1 per reaction. Mix thoroughly by vortexing, centrifuge briefly, and place on ice. On ice, transfer 16.8 μL of the fragmentation master mix to each (31.2 μL) purified ss-cDNA sample from **step 1** for a final reaction volume of 48 μL.

Mix thoroughly by gently vortexing the tube, centrifuge briefly, and incubate for 1 h at 37 °C, then for 2 min at 93 °C, then for at least 2 min at 4 °C. Immediately after the incubation, centrifuge briefly to collect the fragmented ss-cDNA at the bottom of the tube or well.

3. Transfer 45 μL of the fragmented ss-cDNA sample to each tube or well and place on ice.
4. Prepare the labeling master mix composed of 12 μL 5× terminal deoxynucleotidyl transferase (TdT) buffer, 1 μL 5 mM DNA labeling reagent, and 2 μL TdT per reaction. Mix thoroughly by vortexing, centrifuge briefly, and place on ice. On ice, transfer 15 μL of the labeling master mix to each (45 μL) fragmented ss-cDNA sample from **step 3** for a final reaction volume of 60 μL. Mix thoroughly by gently vortexing the tube, centrifuge briefly, and incubate for 1 h at 37 °C, then for 10 min at 70 °C, then for at least 2 min at 4 °C.

3.5.4 Hybridization and Scanning

1. The labeled ss-cDNA (5.2 μg) is mixed with 20× eukaryotic hybridization controls, denatured and hybridized to Human Transcriptome 2.0 ST arrays as recommended in the kit (*see* **Note 12**).
2. After hybridization at 60×*g* for 16 h at 45 °C, the arrays are subjected to a fluidics protocol that washes and stains the array with streptavidin-phycoerythrin.
3. The stained arrays are then scanned in a GeneChip 3000G scanner and the data is exported as CEL files.

3.6 Analysis of Human Transcriptome 2.0 ST Array Data

3.6.1 Generation of CHP Files

1. We have successfully used the Affymetrix Expression Console [9] and Affymetrix Transcriptome Analysis Console [10] software to analyze transcriptome array data. To obtain a robustly confident list of genes associated with a given condition, we use the SST-RMA algorithm as the probe intensity summarization method [6, 11, 12]. The SST-RMA method includes GC4 (GC Correction Version 4) background reduction and SST (Signal Space Transformation) intensity normalization to the expression data processing workflow.
2. We first import the data into Expression Console software that provides an easy way to create summarized expression values (CHP files) for individual files or collections of expression array feature intensity (CEL) files. For gene level analysis, select the Gene-level-SST-RMA workflow and for alternate-splicing analysis select the Exon-level-SST-Alt-Splice-Analysis workflow to create the respective CHP files.

3.6.2 Gene Level Analysis

1. Traditional microarray analysis methods present a steep learning curve for the average user. The problem resides primarily in the normalization techniques used to distribute the signal intensities

on the array. To obtain a robustly confident list of genes associated with a given condition, we use the SST-RMA algorithm as the probe intensity summarization method [6, 11, 12].

2. We strongly recommend the use of replicates in the experiments using microarray technology for gene expression profiling. While we realize that these experiments can be cost prohibitive, confidence in that data from microarray experiments requires the use of at least three biological replicates (with five or more replicates being preferred in the case of biological samples with rare transcripts or high level of variability).
3. After summarization, we routinely conduct a Principal component analysis to identify any outliers in the samples. We also evaluate the control spikes and hybridization metrics as described by Affymetrix [7].
4. The gene level CHP files created in Expression Console are imported into Transcriptome Analysis Console software and analyzed using the Gene Level Differential Expression Analysis workflow.
5. A statistical test (Students *t*-test or ANOVA) with a *p*-value < 0.05 and a false discovery rate correction (Benjamini Hochberg) routine is most appropriate at this step. The stringency of the statistics will determine how many differentially expressed targets are identified.
6. We further reduce the data by applying a filter on fold change of expression values between the two conditions. While a two-fold cutoff seems to be used in many microarray experiments, we prefer to use a 1.5-fold cutoff. This enables us to have enough probe sets in our lists while performing secondary analysis such as gene ontology or pathway analysis.

3.6.3 Transcriptome Splicing Analysis (Exon Level)

1. For exon-level analysis, the alt-splice CHP files created using Expression Console are imported in to Transcriptome Analysis Console software and analyzed using the Alternative Splicing Analysis workflow.
2. This is followed by a splicing ANOVA with a *p*-value < 0.05. This uses a gene-normalized intensity value, i.e., ratio of probe set intensity to expression level of the gene (Table 1).
3. A Splicing Index value is then calculated. This is similar to a fold-change filter where the gene-normalized intensity values are compared between the two experimental conditions. Splicing Index algorithm is a way to measure how much exon-specific expression differs between two conditions after excluding gene level influences. The algorithm first normalizes the exon and junction expression values by the level of gene expression and creates a ratio of normalized signal estimates from one condition relative to another.

Table 1 Splicing analysis of CXCL2 using TAC software reveals an alternative 3′ acceptor site

Transcript cluster ID	Gene fold changes (linear) siR	Gene symbol	Description	Group	PSR/junction ID	Splicing index (linear) (siRNA#12 vs…)	ANOVA *p*-value (siRNA = 12 vs. control…)	FDR *p*-value (siRNA#12 vs. control…)	Splicing event estimate (siRNA#12 vs. control siRNA)
TC04001286…	−6.26	CXCL2	Chemokine (C-X motif) ligand 2; NULL	Coding	PSR04017577.h…	4.34	0.023088	0.337821	Intron retention
					PSR04017573.h…	−6.66	0.000003	0.083603	Alternative 3′ acceptor site
					PSR04017570.h…	7.19	0.007015	0.242672	Alternative 3’ acceptor site
					PSR04017578.h…	−4.31	0.005124	0.224651	Cassette exon
					PSR04017575.h…	2.46	0.015121	0.298744	Alternative 3’ acceptor site
					PSR04017583.h…	3.93	0.002606	0.191474	Alternative 5′ donor site
					PSR04017580.h…	2.23	0.007989	0.251277	Cassette exon
					JUC0400392.h…	−6.18	0.002517	0.190013	
					JUC04009393.h…	2.04	0.046840	0.414915	

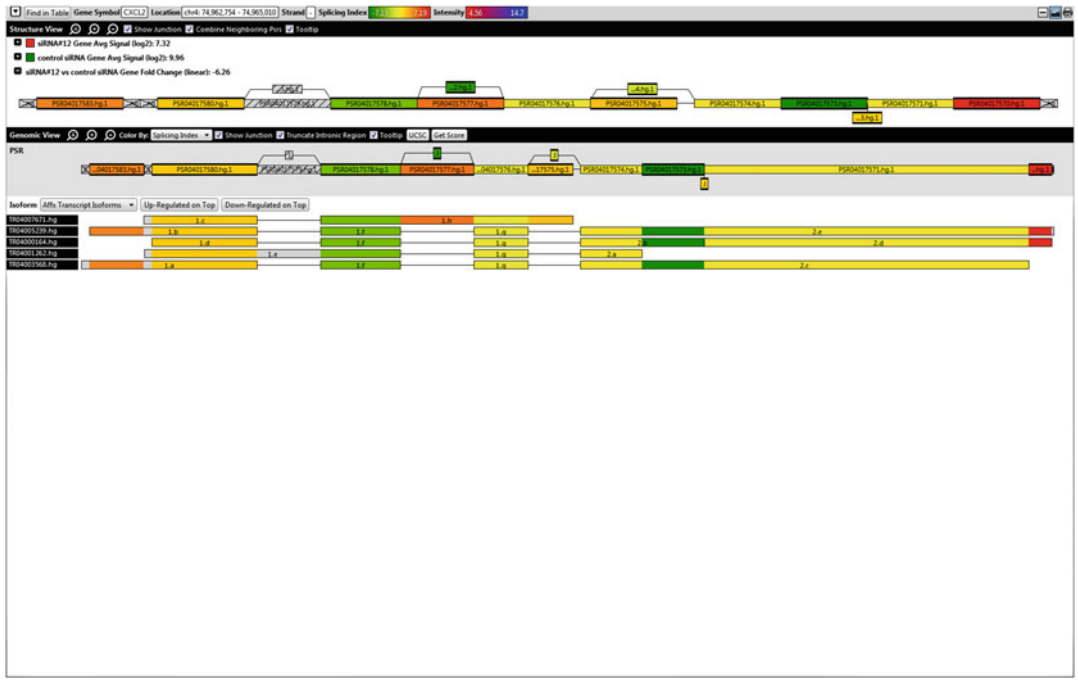

Fig. 2 Alternate splicing events observed in CXCL2 using the visualization tool in TAC software

4. The default criteria for alt-splicing analysis include (a) Splicing Index (linear) < −2 or Splicing Index (linear) > 2, (b) ANOVA *p*-value < 0.05, (c) a gene is expressed in both conditions, (d) a Probe Selection region (PSR)/Junction must be expressed in at least one condition, and that (e) a gene must contain at least one PSR.
5. On average the HTA 2.0 array contains ten probes per exon and four probes per junction. The list of probes that pass the above steps can then be visualized. *See* Fig. 2 for an example of alternative splicing in the 5′end of CXCL2.

   The results of any microarray experiment should be verified using an independent technique such as quantitative PCR or sequencing. Additional functional analysis is also recommended.

4 Notes

1. All instruments, glassware, and plasticware that touch cells or cell lysates should be certified DNase-free and RNase-free or should be prewashed with RNaseZap (Ambion, cat. #9780; 9782) or RNase AWAY (Molecular BioProducts cat. #7001) followed by DEPC water and allowed to air-dry.

2. The number of cells required for each microarray experiment can vary from cell type to cell type. Typically we utilize a 25 or 75 cm^2 flask of confluent cells per condition. This corresponds to about $2–10 \times 10^6$ cells and provides enough material for both the microarray experiment as well as other validation and QC experiments.
3. If using this RNA for any miRNA analysis, avoid the column (unless it is specified for miRNA use) cleanup step since it results in loss of small RNAs.
4. Ambion and Affymetrix protocols and technical literature (and our experience) suggest that samples failing to meet either (or both) of these criteria may (or will) perform poorly in molecular techniques, which are based on reverse transcription followed by amplification. This is likely due to the interference of protein, carbohydrate, or phenolic contaminants on the reverse transcription process.
5. The Agilent BioAnalyzer is a preferable substitute to MOPS-formaldehyde agarose gel analysis due to the reduced sample required, increased sensitivity, and reduced exposure to toxic reagents. The Agilent 2100 Expert software provides a RIN or RNA integrity number for the RNA nano and pico assays (series II). It is recommended that this RIN number be between 7 and 10 if the RNA sample is to be used in a microarray experiment. We generally use the RIN number as a secondary QC criteria along with 260/280, 260/230, and 28S/18S ratios.
6. All the reagents for this protocol are supplied in the Affymetrix WT PLUS Reagents kit [8]. It is recommended that polyA RNA controls be spiked in to the starting RNA samples since this will allow to QC for any degradation occurring during the protocol. The signals from these spikes can also be used for normalization. We have also successfully used the Affymetrix WT Pico kit to generate data from transcriptome arrays.
7. Holding the first-strand cDNA synthesis reaction at 4 °C for longer than 10 min may significantly reduce cRNA yields.
8. Disable the heated lid of the thermal cycler or keep the lid off during the second-strand cDNA synthesis.
9. Transfer the second-strand cDNA samples to room temperature for ≥5 min while preparing IVT Master Mix. After the IVT buffer is thawed completely, leave the IVT buffer at room temperature for ≥10 min before preparing the IVT Master Mix.
10. Preheat the nuclease-free water in a heat block or thermal cycler to 65 °C for at least 10 min. Mix the purification beads thoroughly by vortexing before use to ensure that they are fully dispersed. Transfer the appropriate amount of purification beads to a nuclease-free tube or container, and allow the purification beads to equilibrate at room temperature. For each

reaction, 100 µL plus ~10% overage will be needed. Prepare fresh dilutions of 80% ethanol wash solution each time from 100% ethanol (molecular biology grade or equivalent) and nuclease-free water in a nuclease-free tube or container. For each reaction, 600 µL plus ~10% overage will be needed. Transfer the cRNA sample to room temperature while preparing the purification beads.

11. Do not over-dry the beads as this will reduce the elution efficiency. The bead surface will appear dull and may have surface cracks when it is over-dry.
12. It is recommended that hybridization controls be prepared from a master mix. The signal from the controls (bioB, bioC, bioD, and Cre) can be used to qualitatively compare chips being hybridized over time.

The Affymetrix GeneChip® Human Transcriptome Array 2.0 expands the toolkit for comprehensive examination of the human transcriptome. The HTA 2.0 array creates an ideal platform for discovery through incorporation of on average 109 probes per gene, 10 probes per exon, and 4 probes per splice junction while evaluating all known genes and noncoding transcripts. With our growing appreciation for the divergent functions of alternative splice variants and the importance of noncoding RNAs, utilizing this highly sensitive system will be critical to uncovering novel biology.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G,

Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J, International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921

2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291(5507):1304–1351. doi:10.1126/science.1058040, 291/5507/1304 [pii]
3. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282(5396):2012–2018
4. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A,

Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420(6915):520–562

5. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306(5696):636–640. doi:10.1126/science.1105136, 306/5696/636 [pii]
6. Kornblihtt AR et al (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14(3):153–165
7. Affymetrix Human Transcriptome 2.0 array. http://media.affymetrix.com/support/technical/datasheets/hta_array2_0_datasheet.pdf
8. Affymetrix WT PLUS Reagent Kit User Manual P/N703147rev3. http://media.affymetrix.com/support/downloads/manuals/wtplus_reagentkit_assay_manual.pdf
9. Affymetrix Expression Console v 1.4 Manual P/N 702387 rev 5. http://media.affymetrix.com/support/downloads/manuals/expression_console_userguide.pdf
10. Affymetrix Transcriptome Analysis Console (TAC) 3.0 Manual P/N 703150 rev 4. http://www.affymetrix.com/support/downloads/manuals/tac_user_manual.pdf
11. Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30(1):13–19
12. Ward AJ, Cooper TA (2010) The pathobiology of splicing. J Pathol 220(2):152–163

Chapter 19

Large-Scale RNA Interference Screening to Identify Transcriptional Regulators of a Tumor Suppressor Gene

Matteo Forloni, Thuy Ho, Lisha Sun, and Narendra Wajapeyee

Abstract

RNA interference (RNAi) is a powerful research tool that can be used to silence the expression of a specific gene. In the past several years, RNAi has provided the opportunity to identify factors and pathways involved in complex biological processes by performing unbiased loss-of-function screens on a genome-wide scale. Here we describe a genome-wide RNAi screening strategy to identify factors that regulates epigenetic silencing of a specific tumor suppressor gene, using RASSF1A as an example. The approach we describe is a general RNAi screening strategy that can be applied to identify other factors that drive and/or maintain epigenetic modifications on specific genes, including cancer-related genes.

Key words RNA interference, Epigenetic silencing, Positive-selection screen, RNA interference, siRNA, shRNA

1 Introduction

High-throughput RNAi screening provides the opportunities to identify in unbiased manner cellular genes associated with specific biological phenotypes and has the potential to identify new targets for therapeutic interventions. Several factors must be considered when designing a successful RNAi screen. General guidelines for choosing the appropriate RNAi library (e.g., shRNA vs. siRNA, retroviral vs. lentiviral) and screening strategy (single well vs. pooled format, positive vs. negative selection, etc.) have been covered in detail elsewhere, and the reader is referred to several excellent reviews on these topics [1–4].

When planning a genome-scale RNAi screening for epigenetic factors, there are several critical parameters to consider. First, it is important to clone a suitable promoter region. For example, as hypermethylation usually occurs at DNA segments abundant with CpG dinucleotides, it will be of necessary to clone the whole promoter region carrying these CpG islands. Second, it is important to clone the promoter of the gene of interest in a suitable reporter

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_19,

plasmid, which in turn can be used for the selection of cells that eventually expression the tumor suppressor gene of interest. Third, a suitable cell line must be chosen in which the gene of interest is epigenetically silenced, yet can easily be reexpressed by treatment of the cells with epigenetic agents, such as DNA methyltransferase inhibitors and histone deacetylase inhibitors. Fourth, including appropriate negative control will allow one to evaluate the background of the screen. In most cases, a control non-silencing shRNA or luciferase siRNA is suitable for this purpose. Moreover, in order to discern off-target effects of shRNAs, it is very important to use multiple, unrelated shRNAs/siRNAs targeting the same gene. Finally, the choice of appropriate assays to validate the positive candidates from the RNAi screening is also important to consider because a clear read out will reduce the background and increase the likelihood of success of the RNAi screen.

Here, we describe an RNAi screening that has been published by our group [5] in which we identified factors involved in epigenetic silencing of the tumor suppressor RAS association domain family 1A (RASSF1A). In brief, we generated a reporter construct in which the RASSF1A promoter was used to direct expression of a gene encoding red fluorescent protein (RFP) fused to the blasticidin resistance ($Blast^R$) gene. This RASSF1A-RFP-$Blast^R$ reporter construct was stably transduced into human MDA-MB-231 breast cancer cells in which the endogenous RASSF1A gene is epigenetically silenced [6]. We then selected cells in which the reporter gene had been silenced as evidenced by loss of RFP expression and acquisition of blasticidin sensitivity. We used a human shRNA library [7] comprising ~62,400 shRNAs directed against 28,000 genes. The shRNAs were divided into ten pools, which were packaged into retrovirus particles and used to stably transduce the MDA-MB-231/RASSF1A-RFP-$Blast^R$ reporter cell line. Blasticidin-resistant colonies, indicative of derepression of the epigenetically silenced reporter gene, were selected, and the shRNAs were identified by sequence analysis. One positive candidate was further confirmed. Indeed, stable transduction of parental MDA-MB-231 cell line with a single shRNA directed against the candidate gene led to derepression of the endogenous, epigenetically silenced RASSF1A gene. Confirmed candidate shRNAs were then tested in a secondary screen for their ability to promote derepression of endogenous epigenetically silenced RASSF1A in three independent NSCLC cell lines: A549, NCI-H23, and NCI-H460 [5].

The approach described here is a general screening strategy that can be used to study other epigenetically silenced genes in different human cancer (or mouse) cell lines.

2 Materials

2.1 Generation of the Reporter Construct, Cell Transfection, and Selection of Stable Clones

1. BAC.
2. pDsRed2-N1 (Clontech).
3. PEF6V5-HisB (Invitrogen).
4. Neomycin.
5. Blasticidin.
6. Transfection reagent, such as Effectene (Qiagen).
7. MDA-MB-231 cell line.

2.2 Generation of Retroviral Particles, Determination of the Multiplicity of Infection, and Cell Infection

1. Genome-wide shRNA library and a control shRNA, such as a non-silencing shRNA (Open Biosystems or Sigma-Aldrich).
2. Phoenix-gp helper-free retrovirus producer cell line (Garry Nolan, Stanford University; l).
3. pCI-VSVG plasmid (Addgene).
4. pGag-pol (Addgene).
5. Transfection reagent, such as Effectene.
6. 0.45 mM filters.
7. Culture medium: DMEM high glucose (1×, liquid, with L-glutamine and sodium pyruvate), 10% FBS, and penicillin–streptomycin.
8. Polybrene.
9. Puromycin.
10. Crystal violet staining solution: 40% methanol, 10% acetic acid, 50% ddH_2O, 0.01% crystal violet.
11. MDA-MB-231 cell line.

2.3 Isolation of Genomic DNA and Identification of Candidate shRNAs by DNA Sequencing

1. Trypsin-EDTA (0.25%, Invitrogen).
2. Genomic DNA preparation buffer: 100 mM NaCl, 10 mM Tris–HCl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% (v/w) SDS, 50 mL of proteinase K.
3. Phenol–chloroform–isoamyl alcohol (25:24:1).
4. Chloroform.
5. NaCl (5 M).
6. Ethanol (70 and 100% solutions).
7. TE buffer (1×): 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
8. Spectrophotometer or NanoDrop.
9. 5× Go-Taq PCR buffer.
10. Taq DNA polymerase.

11. Primers for sequencing shRNA inserts in pSM2 library: For-pSM2 (5′-GCTCGCTTCGGCAGCACATATAC-3′) and Rev-pSM2 (5′-GAGACGTGCTACTTCCATTTGTC-3′).
12. DNase- and RNase-free agarose for gel electrophoresis.
13. Ethidium bromide solution (10 mg/mL).
14. QIAquick gel extraction kit.
15. pGEM-T Vector system I (Promega).
16. Bacterial competent cells that allow for blue/white selection, such as DH5a, and that have a transformation efficiency of $>10^6$ colonies/μg.
17. LB-agar plates with 100 μg/mL ampicillin, 40 μL X-gal (50 mg/mL), and 10 μL IPTG (1 M).
18. LB liquid.
19. QIAprep Miniprep kit (Qiagen).
20. SP6 sequencing primer (sequence 5′-ATTTAGGTGACAC TATAG-3′).

3 Methods

3.1 Generation of the RASSF1A-RFP-Blast^R Reporter Construct, Cell Transduction, and Selection of Stable Clones

1. Excise the CMV promoter from pDsRed2-N1.
2. PCR amplify the blasticidin resistance gene from a plasmid, such as PEF6V5-HisB.
3. Clone the blasticidin resistance gene into pDsRed2-N1 to generate an in-frame fusion with DsRed2 gene.
4. Amplify 2.5 Kb of the RASSF1A promoter from a BAC.
5. Clone the RASSF1A promoter fragment into the derivative of DsRed2-N1 with the blasticidin resistance gene.
6. Transfect MDA-MB-231 cells with the RASSF1A-RFP-Blast^R reporter construct by using a transfection reagent, such as Effectene (Qiagen).
7. After 24 h enrich for the stable clones by selecting with neomycin.
8. Use in the subsequent RNAi screen the clones that show epigenetic silencing of the RASSF1A promoter, as observed by blasticidin sensitivity and lack of RFP expression.

3.2 Generation of Retroviral Particles

1. Plate 3×10^6 Phoenix-gp cells in ten individual 100 mm tissue culture dishes. Plate one additional dish, to be infected with a retrovirus expressing a control non-silencing shRNA.
2. After 36 h, transfect cells with 10 μg pooled shRNA plasmid DNA (*see* **Note 1**), 1 μg Gag–pol plasmid DNA, and 1 μg pCI-VSVG plasmid DNA using a transfection reagent.

3. After 48 h, collect the culture supernatants, which contain retroviral particles.
4. Filter the culture supernatants using 0.45 μM filters. Aliquot 1 mL supernatant into microfuge tubes and freeze at −80 °C (*see* **Note 2**).

3.3 Determining the Multiplicity of Infection for Retroviral shRNA Pools

1. Plate 1×10^5 293 cells in each well of a 6-well plate, using one plate for each pool.
2. Perform serial dilutions of each retroviral shRNA pool. First, label six microfuge tubes as "10^{-1}," "10^{-2}," "10^{-3}," "10^{-4}," "10^{-5}," and "10^{-6}." Add 1 mL of the DMEM media to the first tube and add 1000 and 900 μL in the tubes from 2 to 6. Add 100 μL of retroviral supernatant in the first tube, resulting in a 1/10 dilution (10^{-1}). Remove 100 μL of the 10^{-1} dilution and add it to the second tube to create a 10^{-2} dilution. Repeat to generate subsequent serial dilutions.
3. Aspirate the media from all the wells and add 1 mL of serially diluted retroviral supernatant with polybrene (10 μL/mL) to the appropriate well.
4. After 24 h, remove the media and add 2 mL of fresh DMEM media.
5. After 24 h, add puromycin (1.0 μg/mL) to select for cells carrying the retroviral shRNA.
6. Change the media with puromycin every 3 days.
7. Between day 10 and 14, depending upon the size of the colonies, stain the colonies that survive the puromycin selection using crystal violet staining solution.
8. Calculate the multiplicity of infection (MOI) of the retroviral supernatants as follows:

$$\text{MOI}\left(\text{particle forming units}\left(\text{pfu}\right)/\text{mL}\right) = \text{Number of colonies} \times \text{dilution factor} \times 10.$$

For example, if you observe five colonies in the 10^{-4} dilution plate, the calculation will be: $5 \times 10^4 \times 10 = 5 \times 10^5$ pfu/mL (*see* **Note 3**).

3.4 Infection and Selection of Cells After Transduction with Retroviral shRNA Pools

1. Plate 1.2×10^6 MDA-MB-231 cells in ten individual 100 mm tissue culture dishes.
2. After 24 h, transduce the MDA-MB-231 cells with retroviral shRNA pools in a total volume of 5 mL of DMEM media with 10% FBS/penicillin–streptomycin and polybrene (10 μg/mL) to achieve infection at an MOI of 0.2 (*see* **Note 4**).
3. After 24 h, change the media and add 10 mL of DMEM media with 10% FBS/penicillin–streptomycin.

4. After 24 h, add puromycin (1.0 μg/mL) to enrich for cells that carry integrated shRNAs. Change the media every 3 days with fresh puromycin.
5. After 5–7 days, when the puromycin selection is over, add blasticidin (2.0 μg/mL) to select for blasticidin-resistant colonies, indicative of derepression of the epigenetically silenced reporter gene.

3.5 Isolation of Genomic DNA from Blasticidin-Resistant and RFP-Positive Cells and Identification of Integrated shRNAs by DNA Sequencing

1. For all ten pools, trypsinize and isolate the cells that acquired blasticidin resistance.
2. Extract genomic DNA using Qiagen genomic DNA isolation kit as per the manufacturers' instructions.
3. To amplify the retroviral shRNA integrated into the genomic DNA, set up the following PCR:

Components	Volume
5× Go-Taq PCR buffer	10 μL
Taq polymerase (5 units/μL)	Taq polymerase (5 units/μL)
Genomic DNA	2 μL [100 ng (50 ng/μL)]
For-pSM2 (10 pmoles/μL)	1 μL
Rev-pSM2 (10 pmoles/μL)	1 μL
dd H_2O	35 μL

4. Run the PCR products on a 1% agarose gel with 10 μL ethidium bromide (10 mg/mL stock) and elute from the gel using a Qiagen gel elution kit.
5. Ligate 100 ng of the eluted PCR product with the TA vector using the TA cloning kit (Promega) as per the manufacturer's instructions.
6. Perform an overnight ligation at 16 °C.
7. Next day, transform ligation mixture into bacterial competent cells, and plate the reaction onto LB-agar plates containing ampicillin, 40 μL of X-gal (50 mg/mL), and 10 μL of IPTG (1 M).
8. Inoculate white colonies into tubes containing 3 mL of LB liquid with 100 μg/mL ampicillin. Grow overnight at 37 °C.
9. Isolate plasmid DNA from white colonies using Qiagen's miniprep kit as per the manufacturer's instructions.
10. Sequence the plasmid DNA using the SP6 primer.
11. To identify the genes targeted by the shRNAs, perform a nucleotide BLAST search using the shRNA sequence as a query.

3.6 Secondary Assays to Confirm the Candidates Identified from the Initial Screening

1. Select the individual shRNA corresponding to the candidate gene from the RNAi library and prepare the retrovirus as described above in Subheading 3.1.
2. Plate 1×10^5 MDA-MB-231 cells in a well of a 6-well for each candidate shRNA to be tested, and infect with the retrovirus particle(s).
3. After 24 h, change the media and add 2 mL of DMEM media.
4. After 24 h, add puromycin (1.0 μg/mL) to enrich the cells that carry integrated shRNAs. Change the media every 3 days with fresh puromycin.
5. Validate the candidates by using several assays to assess if the knockdown promotes the derepression of the endogenous epigenetically silenced target gene. These assays include qRT-PCR as well as immunoblot.
6. Validate the positive candidate by using other shRNAs, unrelated in sequence to the first shRNA, to rule out "off-target" effects (*see* **Note 5**).
7. To further confirm the candidates, repeat the validation in several cell lines in which the gene of interest is epigenetically silenced.

4 Notes

1. DNA amount along with the packaging plasmids should be optimized to obtain high retroviral titer.
2. It is important to note that freeze-thawing will lead to drop in virus titer Therefore, we recommend avoiding multiple freeze-thaw cycles. Storing viral supernatant in the form of aliquots at −80 °C freezer prevents the drop of viral titer and improves the infection.
3. While performing the titration, it is important to determine the titer in the cell line that will be eventually used for the screen, rather than in an unrelated cell line, because titer can differ depending upon the infectivity of different cancer cell lines.
4. The purpose of infecting the cells at the MOI of 0.2 is to prevent superinfection and to ensure that each cell receives no more than one shRNA, and thus the observed phenotype is driven due to a single-gene knockdown.
5. It is recommended that at least two to three additional shRNAs are used against the validated candidates to ensure the gene-specific effect and to rule out off-target effects. Additionally, when possible an shRNA-resistant cDNA can be used to perform the rescue experiments.

Acknowledgment

We gratefully acknowledge grants from the National Institutes of Health: R21CA197758-01 (N.W.), R21CA191364-01 (N.W.), R21CA195077-01A1 (N.W.), and R01CA200919-01 (N.W.). N.W. is also supported by Research Scholar Grant from American Cancer Society (128347-RSG-15-212-01-TBG) and grants from the Melanoma Research Alliance and the Melanoma Research Foundation.

References

1. Ashworth A, Bernards R (2010) Using functional genetics to understand breast cancer biology. Cold Spring Harb Perspect Biol 2(7): a003327. doi:10.1101/cshperspect.a003327, cshperspect.a003327 [pii]
2. Lizardi PM, Forloni M, Wajapeyee N (2011) Genome-wide approaches for cancer gene discovery. Trends Biotechnol 29(11):558–568. doi:10.1016/j.tibtech.2011.06.003, S0167-7799(11)00105-3 [pii]
3. Mullenders J, Bernards R (2009) Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. Oncogene 28(50):4409–4420. doi:10.1038/onc.2009.295, onc2009295 [pii]
4. Westbrook TF, Stegmeier F, Elledge SJ (2005) Dissecting cancer pathways and vulnerabilities with RNAi. Cold Spring Harb Symp Quant Biol 70:435–444. doi:10.1101/sqb.2005.70.031
5. Palakurthy RK, Wajapeyee N, Santra MK, Gazin C, Lin L, Gobeil S, Green MR (2009) Epigenetic silencing of the RASSF1A tumor suppressor gene through HOXB3-mediated induction of DNMT3B expression. Mol Cell 36(2): 219–230. doi:10.1016/j.molcel.2009.10.009, S1097-2765(09)00744-8 [pii]
6. Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP (2000) Epigenetic inactivation of a RAS association domain family protein from the lung tumour suppressor locus 3p21.3. Nat Genet 25(3):315–319. doi:10.1038/77083
7. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR, Sheth N, Bradshaw J, Burchard J, Kulkarni A, Cavet G, Sachidanandam R, McCombie WR, Cleary MA, Elledge SJ, Hannon GJ (2005) Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet 37(11): 1281–1288. doi:10.1038/ng1650, ng1650 [pii]

Chapter 20

Transcriptional Analysis-Based Integrative Genomics Approach to Identify Tumor-Promoting Metabolic Genes

Romi Gupta and Narendra Wajapeyee

Abstract

Metabolic regulation can play key role in normal and pathological states. In particular in cancer cells, alterations in metabolic pathways can drive the growth and survival of cancer cells. Among these alterations, many occur at the transcriptional level leading to the overexpression of metabolic genes. However, not every metabolically upregulated genes may be necessary for tumor growth. Therefore, functional validation approaches are required to distinguish metabolically overexpressed genes that are necessary for tumor growth versus the ones that are not. One of the experimental approaches to do this is to use the approach of RNA interference to systematically survey the transcriptionally upregulated metabolic genes for their requirement in tumor growth. Here, we describe an integrative genomics approach to identify metabolic genes that are necessary for tumor growth. The approach we describe is a general integrative genomics approach that combines bioinformatics-based identification of overexpressed metabolic genes in cancer patient samples and then uses RNAi-based knockdown approach to identify genes that are necessary for tumor growth.

Key words RNA interference, Metabolism alteration, Gene expression analysis

1 Introduction

It is now well appreciated that cancer cells arise due to not small set of changes in normal cells, rather a series of multiple changes that make cancer cells adaptable to grow and expand in vivo and counteract a variety of stress and inhibitory situations. Collectively, these features are recognized as hallmarks of cancer. From the time of Otto Warburg discovery of aerobic glycolysis in cancer cells, it became clear that cancer cells show alterations in metabolic pathways and these changes seemed more than just passenger alterations [1]. Since the discovery of "Warburg effect," several studies have identified, and implication metabolic alteration in almost all metabolic pathways in cancer cells and their role has been shown in almost all aspects of cancer growth and metastasis as well as therapy response [1]. In most cases candidate-based approaches have been used to identify genes and pathways that are important for metabolic

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2_20,

regulation of tumor growth. However, in most cases these targeted approaches do not allow a large-scale survey of metabolic alterations in cancer cells. Therefore, systematic approaches to identify metabolic alterations necessary for tumor growth are necessary.

Availability of a large number of microarray and RNA-sequencing data from the patient-derived tumor samples and normal control in public domain gives an ample opportunity to identify metabolic alterations in cancer cells. Once metabolic genes that are transcriptionally upregulated in patient-derived cancer cells are identified, this information can be combined with a targeted RNA interference screening approach to identify which of the transcriptionally upregulated genes that are necessary for tumor growth.

Several factors must be considered when analyzing the microarray or RNA-sequencing datasets for patient-derived cancer samples and designing a successful RNAi screening. General guidelines for choosing the appropriate RNAi library (e.g., shRNA vs. siRNA, retroviral vs. lentiviral) and screening strategy (single well vs. pooled format, positive vs. negative selection, etc.) have been covered in detail elsewhere, and the reader is referred to several excellent reviews on these topics [2–5].

When planning a targeted RNAi screening for transcriptionally overexpressed metabolic genes, there are several critical parameters to consider. First, it is really important to choose the right cancer datasets. Preferably, multiple datasets should be analyzed for each cancer type, and genes that are commonly overexpressed should be considered for analysis. Second, at least two and ideally three sequence-independent short-hairpin (shRNA) or small interfering (siRNA) should be used to confirm the observed effect. Third, it is important to choose a cancer cell line in which the gene of interest is overexpressed, similar to that observed in the case of patient samples. Fourth, it is fundamental to choose an appropriate negative control to evaluate the background of the screen. In most cases, a control non-silencing shRNA or luciferase siRNA is suitable. Finally, the choice is also very important, as these will allow the researcher to clearly distinguish between false and true positive.

Here, we describe an integrative genomics approach to identify metabolic drivers of tumor growth. The approach described here is a general screening strategy that could be used to study and identify metabolic driver of tumor growth.

2 Materials

2.1 Analysis of Patient-Derived Cancer Sample Datasets

1. Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/gds) (publically available).
2. ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) (publically available).
3. Oncomine (https://www.oncomine.org/resource/login.html) (license required for detailed dataset analysis).

4. The Cancer Genome Atlas (http://cancergenome.nih.gov/) (request to access is needed).
5. International Cancer Genome Atlas (https://icgc.org/).
6. cBioPortal (http://www.cbioportal.org/).

2.2 Generation of Retroviral Particles, Determination of the Multiplicity of Infection, and Cell Infection

1. Gene-specific shRNAs (preferably TRC library shRNAs cloned in pLKO.1 vector) and a control shRNA, such as a nonspecific shRNA.
2. 293 T cells (American Type Culture Collection).
3. pMD2.G plasmid (Addgene).
4. psPAX2 plasmid (Addgene).
5. Transfection reagent, such as Effectene (Qiagen).
6. 0.45 mM filters.
7. Culture medium: DMEM high glucose (1×, liquid, with L-glutamine and sodium pyruvate), 10% FBS, and penicillin–streptomycin.
8. Puromycin.
9. Cancer cell line of choice, for example, A735 for melanoma-related experiments.

2.3 Isolation of Total RNA- and shRNA-Specific Gene Knockdown Validation by RT-qPCR

1. Trizol reagent (Invitrogen).
2. RNeasy mini columns (Qiagen).
3. Gene-specific primers.
4. ProtoScript First Strand cDNA synthesis kit (New England Biolabs).
5. Power SYBR green mix (Invitrogen).
6. 384-well qPCR plates.
7. Clear qPCR plate sealing film.
8. Quantitative PCR machine.

2.4 Soft-Agar Assay to Measure Tumorigenic Potential of Cancer Cells Expressing Specific shRNAs

1. DMEM high-glucose media with L-glutamine powder for 1 L (Sigma-Aldrich).
2. 100× penicillin and streptomycin.
3. Trypsin-EDTA (0.25%) (Invitrogen).
4. 1× phosphate-buffered saline.
5. Low gelling agarose.
6. Sterile 6-well tissue culture plates.
7. CO_2 incubator.
8. Water bath.
9. Staining solution (0.05% crystal violet, 40% methanol, 10% acetic acid, and 50% double distilled water).

10. De-staining solution (40% methanol, 10% acetic acid, and 50% double distilled water).
11. Inverted microscope.

2.5 Athymic Nude Mice-Based Tumorigenesis Experiments to Determine the Role of Tumor-Promoting Genes Validated from Soft-Agar Assay In Vivo

1. Athymic nude mice NCR sp/sp (Taconics).
2. Insulin syringe with the 23-gauge needle.
3. Alcohol wipe.
4. Clear air BSL2+ level animal facility.
5. BSL2 laminar airflow.

3 Methods

3.1 Cancer Dataset Analysis

Identify the dataset for analysis using one or more of the databases listed in the Subheading 2.1. GEO datasets available at http://www.ncbi.nlm.nih.gov/gds and can be freely downloaded. Most datasets available through GEO can be analyzed using GEO2R. A detailed instruction for using GEO2R is provided at http://www.ncbi.nlm.nih.gov/geo/info/geo2r.html, and a video for using GEO2R is also available through https://www.youtube.com/watch?v=EUPmGWS8ik0. (*see* **Note 1**).

3.2 Generation of Lentiviral Particles

1. Plate 100,000 293 T cells in 12-well tissue culture dishes using DMEM high-glucose media containing 10% fetal bovine serum and penicillin and streptomycin. One well each can be used for gene-specific shRNAs and for non-silencing shRNA.
2. After 36 h of plating, transfect cells with 0.5 μg of shRNA plasmid DNA, 0.5 μg of psPAX2, and 0.25 μg of pMD2.G using Effectene transfection reagent as per the suppliers' information.
3. After 48 h of transfection, collect the culture supernatants, which contain lentiviral particles.
4. Filter the culture supernatants using 0.45 μM filters. Aliquot 0.5 ml supernatant into 1.5 ml microfuge tubes and freeze at −80 °C (*see* **Note 2**).

3.3 Infection and Selection of Cells After Transduction with Lentiviral shRNA Particles

1. Plate 100,000 cancer cells of choice in 12-well plates in 1 ml of culture media recommended for growing the given cancer cells. If you are testing five gene-specific shRNAs, then the total number of wells required for the experiment will be six, which includes one well for the control-nonspecific shRNA.
2. After 24 h, transduce the cells with 100 μl of supernatant containing lentiviral shRNA particles (*see* **Note 3**). For this

purpose, add the 100 µl of lentiviral shRNA particles containing supernatant directly to the media.

3. After 24 h of injection, change the media and add fresh culture media with 10% FBS and penicillin and streptomycin.
4. After 24 h of changing the media, add puromycin (range 0.2 µg/ml to 2.0 2 µg/ml) to enrich for cells that carry integrated shRNAs (*see* **Note 4**). Change the media every 3 days with fresh puromycin. For most cells puromycin selection is complete by 7–10 days. Use uninfected cells as controls to ensure that the selection is complete.

3.4 Validation of Knockdown Efficiency for shRNAs

1. Isolate total RNA from cells expressing individual gene-specific shRNAs as well as nonspecific shRNA using Trizol reagent as per the manufacturers' instructions.
2. Purify the total RNA using the Qiagen RNAeasy mini column as per the manufacturers' instructions.
3. Perform cDNA synthesis using ProtoScript First Strand cDNA synthesis kit as per the manufacturers' instructions. Dilute the cDNA to two times after the reaction is over. For example, a 20 µl reaction should be diluted to 40 µl.
4. To determine the knockdown efficiency of shRNAs, perform the quantitative PCR. A typical reaction in a 384-well qPCR plate will be as following:

Components	Volume
2× Power SYBR green mix	5 µl
cDNA	2 µl
Gene-specific primer mix	1 µl
Double distilled H_2O	2 µl

5. The relative fold changes in mRNA expression can be calculated using the comparative C_T method [6]. The amount of target gene expression under each indicated condition should be normalized to the expression of human or mouse *GAPDH* as necessary. Relative gene expression among treatment conditions should be calculated using the formula: $-2^{\Delta\Delta C}T$. PCR efficiency of target genes was matched to the PCR efficiency of *GAPDH* by ensuring that log input versus ΔC_T had a slope of zero.

3.5 Soft-Agar Assay to Determine the Tumorigenic Potential of Specific shRNA-Expressing Cancer Cells

1. Select the cells with shRNAs that shows strong knockdown efficiency. Preferentially shRNAs having 50% or more gene knockdowns should be used. Also, include nonspecific shRNA-expressing cells. Use cells that are from 50 to 70% confluency for soft-agar assay.

2. As a first step generate the base agar layer of 0.8% agar. To do this mix equal volume of 1.6% low gelling agar in water, and mix it with 2× DMEM media containing 2× 10% FBS and 2× penicillin–streptomycin. Make sure to mix the 1.6% agarose at no higher than 45 °C to avoid degrading the heat-labile component in culture media. Pour 2 ml/well of the agarose and culture media mix to the 6-well plate.
3. Allow the base agar layer to solidify. Typically it will take about 1 h for base agar layer to solidify.
4. Once the base agar layer solidified, plate the cells (typically in the range of 2500–10,000 cells/well depending upon the cell lines) in 0.4% agarose. To do so, mix 750 μl of 0.8% low gelling agarose with 750 μl of 1× DMEM with 10% FBS and 1× penicillin–streptomycin containing desired number of cells.
5. Allow the top agar to solidify for 30 min and then transfer the plates to the CO_2 incubators.
6. After 24 h of plating the cells, add 1 ml of cell culture media and continue to feed the cells with 1 ml media every 3 days. Typically, it takes 2–3 weeks to obtain right size colonies in soft-agar assay.
7. At the end of experiments, stain the soft-agar colonies with staining solution (0.05% crystal violet solution in 40% methanol, 10% acetic acid, and 50% water) for 2 h. Remove the staining solution and de-stain the plate using de-staining solution (40% methanol, 10% acetic acid, and 50% water) until the colonies start to become visible.
8. Scan and count the number of colonies and also acquire images at 4× and 10× resolution using inverted microscope. The less number of colonies or smaller number of colonies in a soft-agar assay upon the knockdown of a given gene is reflective of the role of a gene in facilitating tumor growth. A secondary assay, potentially an in vivo tumorigenesis assay should be used to confirm this role. Below, we describe how to perform a typical in vivo tumorigenesis assay.

3.6 Athymic Nude Mice-Based Tumorigenesis Experiments to Determine the Role of Tumor-Promoting Genes Validated from Soft-Agar Assay In Vivo

As a secondary validation for the soft-agar results, we recommend to perform the human cancer cell line xenograft-based in vivo tumorigenesis assay using immunocompromised mice. This can be achieved as below:

1. Obtain athymic nude mice from a commercial supplier approved by your institute to avoid long delays in quarantine. For examples, most places direct supplies from Taconic or Jackson Laboratories. At least five mice per group should be used for obtaining statistically significant result. For a more accurate estimate for the animal number, a power analysis can be performed with the potential effect size included in the

analysis. A typical power analysis can be performed using the following website: http://powerandsamplesize.com/.

2. Grow the cells in required amount. Typically five to ten million cells are injected per mice subcutaneously (*see* **Note 5**).
3. Label the mice with ear tags for easy identification.
4. Wipe the flank of athymic nude mice with ethanol wipes, and then using insulin syringe with attached needles, inject five to ten million cells expressing gene-specific shRNAs subcutaneously into the flank of athymic nude mice. As a control, a same number of cells expressing control shRNAs need to be injected in mice, and tumor growth should be compared with the gene-specific knockdown shRNAs.
5. Allow the subcutaneously inject cells to form a tumor and monitor the tumor growth. Typically tumor growth is monitored for the period of 2–3 months depending upon the growth rate of tumor cells.
6. Measure the tumor growth every 3 days using vernier caliper, and calculate the tumor volume using the formula: $\text{length} \times \text{width}^2 \times 0.5$. Reduced tumor growth in vivo in mice due to a specific gene knockdown shows the role of the gene in tumor growth.

4 Notes

1. It is recommended that multiple gene expression datasets are analyzed for each cancer types and this allows data and patient sample collection biases. Also, it is recommended to focus on genes that are upregulated in multiple gene expression datasets compared to the ones that are present only in one.
2. It is important to note that freeze-thawing will lead to drop in virus titer Therefore, we recommend avoiding multiple freeze-thaw cycles. Storing viral supernatant in the form of aliquots at −80 °C freezer prevents the drop of viral titer and improves the infection.
3. Depending upon the titer of lentiviral supernatant, a smaller or larger amount of supernatant can be used.
4. It is recommended that a kill curve is performed using different concentrations of puromycin ranging from 0.1 to 2.0 microgram/ml puromycin to determine the right puromycin concentrations for a given cancer cell line for the selection.
5. In some cases human cancer cell lines may not form tumor in athymic nude mice. In such cases injecting the cells with Matrigel might help. Additionally, using severe combined immunodeficiency (NOD-SCID) mouse model may increase the chance of tumor formation.

Acknowledgment

We gratefully acknowledge grants from the National Institutes of Health: R21CA197758-01 (N.W.), R21CA191364-01 (N.W.), R21CA195077-01A1 (N.W.), and R01CA200919-01 (N.W.). N.W. is also supported by Research Scholar Grant from American Cancer Society (128347-RSG-15-212-01-TBG) and grants from the Melanoma Research Alliance and the Melanoma Research Foundation.

References

1. Pavlova NN, Thompson CB (2016) The emerging hallmarks of cancer metabolism. Cell Metab 23(1):27–47. doi:10.1016/j.cmet.2015.12.006, S1550-4131(15)00621-X [pii]
2. Ashworth A, Bernards R (2010) Using functional genetics to understand breast cancer biology. Cold Spring Harb Perspect Biol 2(7):a003327. doi:10.1101/cshperspect.a003327, cshperspect.a003327 [pii]
3. Mullenders J, Bernards R (2009) Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. Oncogene 28(50):4409–4420. doi:10.1038/onc.2009.295, onc2009295 [pii]
4. Lizardi PM, Forloni M, Wajapeyee N (2011) Genome-wide approaches for cancer gene discovery. Trends Biotechnol 29(11):558–568. doi:10.1016/j.tibtech.2011.06.003, S0167-7799(11)00105-3 [pii]
5. Westbrook TF, Stegmeier F, Elledge SJ (2005) Dissecting cancer pathways and vulnerabilities with RNAi. Cold Spring Harb Symp Quant Biol 70:435–444. doi:10.1101/sqb.2005.70.031
6. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3(6):1101–1108

Index

Narendra Wajapeyee and Romi Gupta (eds.), *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation*, Methods in Molecular Biology, vol. 1507, DOI 10.1007/978-1-4939-6518-2, © Springer Science+Business Media New York 2017

N

P

R

S

T

U

X

Y

Z