Methods in Molecular Biology 1358

Springer Protocols

Erik Dassi *Editor*

Post-Transcriptional Gene Regulation

Second Edition

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

Post-Transcriptional Gene Regulation

Second Edition

Edited by

Erik Dassi

Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Mattarello, Trento, Italy

Humana Press

Editor
Erik Dassi
Laboratory of Translational Genomics
Centre for Integrative Biology
University of Trento
Mattarello, Trento, Italy

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3066-1 ISBN 978-1-4939-3067-8 (eBook)
DOI 10.1007/978-1-4939-3067-8

Library of Congress Control Number: 2015945108

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

The field of post-transcriptional regulation of gene expression (PTR) has been revolutionized in the last few years by the advent of high-throughput techniques empowered by next-generation sequencing, such as ribosome footprinting and protein-RNA interaction determination methods such as CLIP and its variants. This volume is focused on presenting the most recent advances in techniques for studying this important level of gene expression regulation: from bioinformatics approaches, expression profiling, and protein-RNA interactions to noncoding RNAs, RNA modifications, and other aspects, it aims at guiding molecular biologists to harness the power of this new generation of techniques, while also introducing to the data analysis needs these bring along. This book of the *Methods in Molecular Biology* series is organized in six parts: first of all, Part I presents bioinformatics approaches for studying post-transcriptional regulation (Chapters 1 and 2); readers are then introduced to the various expression profiling approaches in Part II (Chapters 3–7). Parts III and IV present protein-RNA interaction and noncoding RNA study techniques (Chapters 8–14). Eventually, Parts V and VI present emerging methods for profiling RNA modifications and other techniques such as alternative translation initiation or polyadenylation sites determination (Chapters 15–22). Recognizing the increasing contribution of bioinformatics in enabling the use of these techniques, several chapters are sprinkled with hints for data analysis, alongside the much needed tips for bench work: the final aim of this volume is thus to offer a versatile resource to the researchers studying post-transcriptional regulation, both introducing the most recent techniques and providing a comprehensive guide to their implementation.

Part I: Bioinformatics

Given the ever increasing amounts of data generated by high-throughput techniques, and the possibility to more easily drive and select the experimental work to be performed through analyses and predictions, ultimately enabling the reduction of the hypothesis space to be explored, bioinformatics is more and more regarded as an invaluable tool for any laboratory studying post-transcriptional regulation of gene expression. Recognizing this essential contribution, the first part thus includes two chapters aimed at illustrating the possibilities offered by current bioinformatics approaches for PTR. Chapter 1 introduces tools and databases dealing with aspects ranging from processing PTR omics datasets to current knowledge about regulatory factors and their interactions with the mRNA, interaction prediction and motif search: such an introduction should allow researchers to start tracing pipelines suited to their analysis needs. Chapter 2, by Marchese and colleagues, instead focuses on a crucial problem for PTR, that of accurately predicting binding sites for RNA-binding proteins (RBPs): the authors thus present their suite of algorithms dealing with this problem, called catRAPID. Eventually, a further bioinformatics chapter (Chapter 12), related in particular to data analysis for CLIP approaches, found its natural place alongside the chapters describing these techniques, and is thus included in Part III.

Part II: Expression Studies

Understanding changes in the output (such as protein synthesis levels, alternative splicing isoforms balance, mRNA degradation) produced by post-transcriptional regulatory mechanisms due to the impact of a stimuli, a treatment or, more in general, to the difference between two conditions is one of the main avenues of research in PTR: indeed, by means of such studies one can formulate hypotheses on the acting trans-factors and related regulatory networks. This part thus aims at presenting tools and techniques allowing to study this aspect, both on a genome-wide and on a smaller scale. First of all, Chapter 3, by Maehr and colleagues, presents a recently introduced technique to perform transcriptional regulation through the Crispr/Cas9 system, empowering the study of PTR changes induced by the absence of a trans-factor of interest. Chapter 4, by Zuccotti and Modelska, introduces us to the polysomal profiling technique, allowing to identify polysomes-bound mRNAs and study trans-factors mediating this association, while in Chapter 5 Spealman and colleagues describe how ribosome positioning on translating transcripts can be studied by means of ribosome profiling, thus allowing translation mechanics in different systems and under various conditions to be systematically analyzed. In Chapter 6, Martinez-Nunez and Sanford present a high-throughput variation of the polysomal profiling technique, tuned for the study of splicing isoform-specific recruitment to the polyribosomes. Eventually, this part is concluded by Chapter 7, in which Chaudhury and Neilson describe a novel reporter system for the analysis of 3′UTR-mediated post-transcriptional gene regulation, which can be used both in vitro and in vivo.

Part III: RBP Interactomics

RNA-binding proteins are one of the biggest players in controlling the post-transcriptional fate of RNA. However, the lack of a reasonably complete catalog of RBP targets is hampering our ability to reconstruct the PTR networks concurring to shape the cell phenotype. Fortunately, this issue has recently received a great deal of attention, and techniques have been developed to tackle the identification of RBPs binding preferences, mode of action (e.g., controlling mRNA stability rather than its splicing), and processes and functions affected by the regulatory activity they exert. This part therefore presents the most recent techniques approaching this task. Chapter 8, by Castello and colleagues, presents the interactome capture method, allowing to identify the repertoire of active RBPs in cultured cells through protein-RNA complexes purification coupled to quantitative mass spectrometry. Chapter 9, by Wessels and colleagues, describes the RIP-seq technique, allowing to immunoprecipitate ribonucleoprotein complexes (RNPs) and identify targets for the RBP of interest in a simple way. Chapter 10, by Danan and colleagues, presents PAR-CLIP, coupling UV-crosslinking and photoactivatable ribonucleosides to immunoprecipitation to obtain precisely defined binding sites for the RBP of interest on its RNA targets; on the same line, Chapter 11 by Sutandy and colleagues describes the iCLIP method, able to generate information about RBP-RNA interactions at single nucleotide resolution. Eventually, as these techniques require specific data analysis toolkits, Chapter 12 by Marvin Jens introduces the reader to a pipeline for PAR-CLIP data analysis, providing a practical usage example and the open-source software implementing it.

Part IV: Noncoding RNAs Interactomics

As identifying RBP targets is of paramount importance for obtaining a complete understanding of PTR networks and mechanisms, so is the determination of noncoding RNAs interactions with mRNAs and other types of RNAs. MicroRNAs (miRNAs), in particular, have emerged as powerful players in many PTR processes, involved both in normal physiology and diseases such as cancer, cardiovascular diseases, and neurodevelopmental syndromes; furthermore, regulatory circuits in which these noncoding RNAs compete with RBP for target regulation have also been described, adding to the complexity of PTR networks. This part thus introduces two techniques tackling miRNA target identification from two different points of view: Chapter 13, by Tan and Lieberman, describes a method for the identification of RNAs bound to a specific miRNA by means of a pulldown and RNA-seq approach, thus providing a comprehensive overview of its function. Adopting a genome-wide perspective, Chapter 14 by Helwak and Tollervey instead presents the CLASH technique, focused on the identification of miRNA-RNA interactions by cross-linking and Argonaute-RNA complexes immunoprecipitation, also recovering the interaction site position through chimeric RNAs formation.

Part V: RNA Modifications

A recently emerging aspect of PTR deals with the many types of post-transcriptional modifications to which an RNA can be subjected. These modifications, ranging from the A-to-I editing to m5C and m6A RNA methylation, have been lately observed to be much more pervasive than previously thought; while many of these affect noncoding RNA species such as tRNAs, mRNAs are now observed to be considerably modified as well. Furthermore, functions and processes affected by these modifications, and their impact on RNA stability, secondary structure, and translation are just beginning to be elucidated. The three chapters composing this part thus aim at presenting techniques to profile the most studied of these modifications: Chapter 15, by Savva and colleagues, thus describes a genome-wide, high signal-to-noise, method tackling the challenging problem of detecting Adenosine (A)-to-Inosine (I) RNA editing sites with high sensitivity and specificity, exploiting single-molecule sequencing. Chapter 16, by Sibbritt and colleagues, presents a technique for the positional profiling of the 5-methylcytosine RNA modification, based on RNA bisulfite conversion and locus-specific, semi-quantitative PCR-based detection of non-converted sites; eventually, Chapter 17 by Liu and Pan concludes this part by describing SCARLET, a technique for the identification and quantitation of N^6-methyladenosine RNA modification sites in mRNAs and long noncoding RNAs at single nucleotide resolution.

Part VI: Other Aspects of PTR

In addition to the PTR facets to which the previous parts have been dedicated, many more processes for which experimental techniques are available take place in the cell and have been at least partially characterized: among these are alternative polyadenylation and translation initiation, splicing, mRNA degradation, and many others; of course, not all of them can be addressed here. This part thus concludes the book by presenting a selection of the

most recent techniques dealing with some of these aspects in PTR. Chapter 18, by Wilkening and colleagues, describes 3′T-fill, a method aimed at the genome-wide identification of alternative polyadenylation sites, thus allowing to profile the impact these have on the post-transcriptional fate of the mRNAs; Chapter 19, by Gao and colleagues, deals instead with the genome-wide identification of alternative translation initiation sites by means of two related approaches, based on ribosome profiling and called GTI-seq and QTI-seq. Chapter 20, by Geisberg and Moqtaderi, presents a technique to profile the half-lives of 3′ mRNA isoforms on a genome-wide level, thus enabling the detailed study of mRNA stability determinants. The last two chapters focus on employing imaging tools to investigate processes related to mRNA metabolism: Chapter 21, by Park and Song, describes a method, based on the MS2-GFP system, for imaging mRNA dynamics, thus allowing to study transport and localization of these molecules in live neurons and brain tissues; eventually, Chapter 22 by Martin and colleagues presents an approach, based on fluorescent tags, to visualize single nascent pre-mRNA molecules and to perform real-time measurement of intron synthesis and excision dynamics.

Finally, I would like to thank all the authors for their invaluable contribution to shaping this book as a passionate and hopefully useful guide to current procedures of this still rising field. Ultimately, it is the fascination for the things we still don't know that leads us to share our knowledge, building on top of it in a scientific feedback loop nourishing our passion.

Mattarello, Trento, Italy *Erik Dassi*

Contents

Contributors

Maria Carmo-Fonseca • *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal*

Alfredo Castello • *Department of Biochemistry, University of Oxford, Oxford, UK*

Arindam Chaudhury • *Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX, USA; Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA*

Susan J. Clark • *Genomics and Epigenetics Department, Garvan Institute of Medical Research, Darlinghurst (Sydney), NSW, Australia*

Charles Danan • *Laboratory of Muscle Stem Cells and Gene Regulation, NIAMS, Bethesda, MD, USA*

Erik Dassi • *Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Mattarello, Trento, Italy*

Katrin Eichelbaum • *European Molecular Biology Laboratory, Heidelberg, Germany*

Bernd Fischer • *German Cancer Research Center (DKFZ), Heidelberg, Germany*

Xiangwei Gao • *Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA*

Joseph V. Geisberg • *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA*

Markus Hafner • *Laboratory of Muscle Stem Cells and Gene Regulation, NIAMS, Bethesda, MD, USA*

Aleksandra Helwak • *Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh, UK*

Matthias W. Hentze • *European Molecular Biology Laboratory, Heidelberg, Germany*

Andrea Hildebrandt • *Institute of Molecular Biology gGmbH, Mainz, Germany*

Antje Hirsekorn • *Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany*

Rastislav Horos • *European Molecular Biology Laboratory, Heidelberg, Germany*

Marvin Jens • *Systems Biology of Gene Regulatory Elements, Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany*

Ana C. de Jesus • *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal*

Nicola A. Kearns • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*

Carl Kingsford • *Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA*

Julian König • *Institute of Molecular Biology gGmbH, Mainz, Germany*

Jeroen Krijgsveld • *European Molecular Biology Laboratory, Heidelberg, Germany*

Judy Lieberman • *Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA*

Nian Liu • *Department of Chemistry, University of Chicago, Chicago, IL, USA*

Carmen Maria Livi • *Centre for Genomic Regulation (CRG), Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain*
René Maehr • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Sudhir Manickavel • *Laboratory of Muscle Stem Cells and Gene Regulation, NIAMS, Bethesda, MD, USA*
Domenica Marchese • *Centre for Genomic Regulation (CRG), Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain*
Robert M. Martin • *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal*
Rocio T. Martinez-Nunez • *Clinical and Experimental Sciences, Sir Henry Wellcome Laboratories, University of Southampton School of Medicine, Southampton General Hospital, Southampton, UK*
Gemma May • *Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA*
C. Joel McManus • *Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA*
Angelika Modelska • *Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Trento, Italy*
Zarmik Moqtaderi • *Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA*
Neelanjan Mukherjee • *Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany*
Joel R. Neilson • *Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX, USA; Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA*
Uwe Ohler • *Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany*
Tao Pan • *Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL, USA, Institute of Biophysical DynamicsUniversity of Chicago, Chicago, IL, USA*
Hye Yoon Park • *Department of Physics and Astronomy, Seoul National University, Seoul, South Korea; The Institute of Molecular Biology and Genetics, Seoul National University, Seoul, South Korea*
Vicent Pelechano • *Genome Biology Unit, EMBL Heidelberg, Heidelberg, Germany*
Hannah Pham • *Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA, USA*
Thomas Preiss • *Genome Biology Department, The John Curtin School of Medical Research (JCSMR), The Australian National University, Acton (Canberra), ACT, Australia, Victor Chang Cardiac Research Institute, Darlinghurst (Sydney), NSW, Australia*
Shu-Bing Qian • *Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA*
Alessandro Quattrone • *Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Trento, Italy*
Robert A. Reenan • *Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA*

José Rino • *Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal*
Jeremy R. Sanford • *Department of Molecular, Cellular and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA, USA*
Yiannis A. Savva • *Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA*
Andrew Shafik • *Genome Biology Department, The John Curtin School of Medical Research (JCSMR), The Australian National University, Acton (Canberra), ACT, Australia*
Tennille Sibbritt • *Genome Biology Department, The John Curtin School of Medical Research (JCSMR), The Australian National University, Acton (Canberra), ACT, Australia*
Minho Song • *The Institute of Molecular Biology and Genetics, Seoul National University, Seoul, South Korea*
Pieter Spealman • *Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA*
Georges St. Laurent • *Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA*
Lars M. Steinmetz • *Genome Biology Unit, EMBL Heidelberg, Heidelberg, Germany; Stanford Genome Technology Center, Stanford University, Palo Alto, CA, USA; Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA*
Claudia Strein • *German Cancer Research Center (DKFZ), Heidelberg, Germany*
FX Reymond Sutandy • *Institute of Molecular Biology gGmbH, Mainz, Germany*
Shen Mynn Tan • *Department of Pediatrics, Boston Children's Hospital, Boston, MA USA*
Gian Gaetano Tartaglia • *Centre for Genomic Regulation (CRG), Barcelona, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain*
David Tollervey • *Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh, UK*
Ji Wan • *Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA*
Hao Wang • *Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA*
Hans-Herman Wessels • *Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany*
Stefan Wilkening • *National Center for Tumor Diseases, Heidelberg, Germany*
Paola Zuccotti • *Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Trento, Italy*

Part I

Bioinformatics

Chapter 1

Introduction to Bioinformatics Resources for Post-transcriptional Regulation of Gene Expression

Alessandro Quattrone and Erik Dassi

Abstract

Untranslated regions (UTRs) and, to a lesser extent, coding sequences of mRNAs are involved in defining the fate of the mature transcripts through the modulation of three primary control processes, mRNA localization, degradation and translation; the action of trans-factors such as RNA-binding proteins (RBPs) and noncoding RNAs (ncRNAs) combined with the presence of defined sequence and structural cis-elements ultimately determines translation levels. Identifying functional regions in UTRs and uncovering post-transcriptional regulators acting upon these regions is thus of paramount importance to understand the spectrum of regulatory possibilities for any given mRNA. This tasks can now be approached computationally, to reduce the space of testable hypotheses and to drive experimental validation.

This chapter focuses on presenting databases and tools allowing to study the various aspects of post-transcriptional regulation, including motif search (sequence and secondary structure), prediction of regulatory networks (e.g., RBP and ncRNA binding sites), profiling of the mRNAs translational state, and other aspects of this level of gene expression regulation. Two analysis pipelines are also presented as practical examples of how the described tools could be integrated and effectively employed.

Key words Bioinformatics, UTR, Database, Prediction, Data analysis, Pipeline, Omics, Polysomal profiling, RBP, ncRNA, Binding site, Secondary structure, Motif

1 Introduction

Post-transcriptional regulation of gene expression (PTR) has been object, in recent years, of an ever increasing interest leading to the development, also thanks to the advent of high-throughput techniques such as microarrays and next-generation sequencing, of a whole new set of experimental assays aiming at profiling and unraveling these mechanisms on a genome-wide scale [1–3]. The advent of next-generation sequencing has indeed provided the possibility to address tasks that were previously out of reach, such as determining an RBP binding specificity: however, in order to fully exploit the huge amounts of data produced by such experiments, one needs a well-built toolkit of analysis tools and databases, which

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_1,

should ultimately allow to trace the regulatory networks underlying the system and conditions under study, thus deriving meaningful functional insights.

A common example of task one could perform is to profile and quantify PTR under different stimuli or treatments, by genome-wide profiling of translating transcripts: this is usually addressed with ribosome or polysome profiling techniques [2, 4]. The resulting datasets need, first of all, specific tools to properly identify differences in translated and untranslated transcripts sets; another class of tools, regulatory elements and binding sites databases/predictors, are then needed to study the UTRs of the thus identified interesting gene sets, trying to understand which specific trans-factors among RBPs and ncRNAs may be playing a role in producing the observed differences, and thus in uncovering related regulatory mechanisms.

More in general, one may need to study PTR events, as mediated by UTRs (or, more rarely, coding sequences) on a single gene or on a set of interesting ones, even if they have not been prioritized after genome-wide analyses. Sequence and structure features of such UTRs must thus be investigated by exploiting existing knowledge (by means of databases) and prediction algorithms capabilities.

Another example of a frequent task, from an opposite perspective, consists in determining binding specificities for an RBP: this is usually approached by the CLIP family of techniques [5–7], by CRAC/CLASH [8] or by methods such as RNAcompete [9], Bind-n-Seq [10], and SEQRS [11]. Data analysis for such assays is far from trivial, and requires dedicated tools and statistical approaches (as Chapter 12 of this book presents in detail). Results (i.e., identified RNA targets and binding sites) will then be amenable to further analyses, such as sequence and secondary structure motif search, using yet another class of tools in order to eventually entirely describe the RBP specificity.

These two tasks we just briefly described are of course only two representative examples of the many possible data analysis workflows one may need to implement while studying PTR; however, they give a clear indication that, in order to be effective in these data analysis tasks, one should invest time in learning about available tools, what they are best suited for and how to combine them.

This chapter first introduces the reader to analysis tools, databases, prediction algorithms, and other software which are currently available to address such tasks, with a particular emphasis on practical considerations and integration of multiple resources: by means of two example pipelines, we then aim at providing guidance in setting up and performing a complete analysis of typical PTR datasets.

2 Tools for Omics Datasets Analysis

The first class of tools we describe deals with omics datasets: these resources, listed in Table 1, can be divided in three broad groups according to the kind of data they deal with. *Expression profiling* datasets are derived by polysome and ribosome profiling followed by RNA-seq and allow us to quantify the mRNAs translational efficiency; *small RNA profiling* datasets are obtained by RNA-seq performed on the short RNAs and permit to identify classes such as miRNAs and piRNAs; eventually, *binding* datasets aim at identifying targets and target binding sites for an RBP and can be produced by techniques such as RIP, CLIP, and their variants. We now proceed to describe and compare all tools related to these experimental approaches.

2.1 Expression Profiling

Polysomal profiling consists in the isolation of polysome-bound, and thus actively translating, transcripts by means of a sucrose gradient separation; this assay is usually complemented with a total mRNA profiling of the same sample, in order to compute translational efficiency (TE) and identify regulatory events affecting this value. **anota** [12] is an algorithm, available as an R package, stemming from the consideration that polysomal mRNA levels are dependent on the total (cytoplasmic) mRNA amounts in the samples. The method thus performs an analysis of partial variance in combination with a random variance model, to avoid considering false buffering events, produced by the commonly used log-ratio approach, as true differential translation phenomena. **tRanslatome** [13] is also an R package, but adopting a wider perspective: it allows the user to identify differentially expressed genes with several methods (limma, *t*-test, TE, RankProd, anota, DEseq, and edgeR), and to perform Gene Ontology and regulatory enrichment analyses; every step is then illustrated through a variety of plot types (scatterplot, MA, SD and identity plot, histogram, heatmap, similarity and radar plot).

Ribosome profiling, a much more recent technique, is instead based on a nuclease protection assay and provides a snapshot of ribosomal 'footprints' on the mRNAs, allowing a quantification of their translational levels; to analyze this kind of data, **Babel** [14] was proposed as a statistical framework to assess the significance of translational control differences between conditions. Available as an R package, it infers an expected ribosome occupancy level (based on mRNA expression) by means of an errors-in-variables regression model; significance of the deviation of the actual ribosome occupancy from this estimate is then assessed by a parametric bootstrap procedure, thus obtaining genes *p*-values.

Table 1

Data analysis tools for PTR omics datasets

Name	Application	Supported data types	Approach	URL	Ref. No.
Anota	Expression	Polysomal profiling	Analysis of partial variance tuned to avoid "false buffering" associated to cytosolic mRNA levels correction	bioconductor.org/packages/release/bioc/html/anota.html	[12]
Babel	Expression	Ribo-seq	Based on an errors-in-variables regression model to identify translational control changes	cran.r-project.org/web/packages/babel/	[14]
tRanslatome	expression	Polysomal profiling	Implements several DEGs identification methods, plot types and functional analyses	bioconductor.org/packages/release/bioc/html/tRanslatome.html	[13]
CPSS	Small RNA	RNA-seq	Web server integrating available analysis tools for small RNA prediction and quantification	mcg.ustc.edu.cn/db/cpss	[15]
iMir	Small RNA	RNA-seq	Pipeline performing miRNA prediction and quantification by several available tools	labmedmolge.unisa.it/inglese/research/imir	[16]
UEA sRNA workbench	Small RNA	RNA-seq	Pipeline performing small RNA prediction, quantification and visualization by ad hoc implemented tools	srna-workbench.cmp.uea.ac.uk	[17]

PIRANHA	Binding	RIP-seq, HITS-CLIP, PAR-CLIP, iCLIP	Allows external covariates, such as expression data, to guide binding sites identification	smithlab.usc.edu	[18]
ASPeak	Binding	RIP-seq, HITS-CLIP	Computes expression-sensitive backgrounds to improve peak calling	sourceforge.net/projects/as-peak	[19]
RIPSeeker	Binding	RIP-seq, PAR-CLIP	Infers significant binding regions by EM and HMM modeling	bioconductor.org/packages/release/bioc/html/RIPSeeker.html	[20]
PARalyzer	Binding	PAR-CLIP	Employs a nonparametric kernel-density estimate classifier to identify binding sites from T>C conversions and read density	genome.duke.edu/labs/ohler/research/PARalyzer	[24]
PIPE-CLIP	Binding	HITS-CLIP, PAR-CLIP, iCLIP	Customizable Galaxy pipeline based on zero-truncated negative binomial likelihoods	pipeclip.qbrc.org	[21]
hyb	Binding	CRAC/CLASH, HITS-CLIP	Complete pipeline allowing chimeric reads detection and folding	https://github.com/gkudla/hyb	[23]
wavClusteR	Binding	PAR-CLIP	Uses a mixture model and a wavelet-based peak calling procedure exploiting coverage function geometry at binding sites regions	bioconductor.org/packages/devel/bioc/html/wavClusteR.html	[25]

The table lists tools available to process PTR omics datasets: listed are the application of each tool (either expression, i.e., translation levels determination, small RNA, i.e., miRNA and other ncRNAs profiling, or binding, i.e., RBP targets identification), the type of supported assays, their specific approach and where the software can be obtained from (URL and reference)

2.2 Small RNA Profiling

Identification and quantification of short RNAs such as, for instance, miRNAs and piwi-interacting RNAs (piRNAs), is performed by means of an RNA-seq assay in which the RNAs selected for sequencing are approximately 20–30 nucleotides long. **CPSS** [15] is a freely available webserver combining existing tools to analyze NGS smallRNA data, including quantification, differential expression, target predictions, functional analysis and novel miRNA identification: the output is presented as a convenient graphic summary in the browser, with detailed results as downloadable files; on the same line, also **iMir** [16] is a pipeline integrating many preexisting open source tools and providing an easy to use graphical user interface to analyze NGS data for the identification of small ncRNAs (existing and novel), for the analysis of their differential expression and for prediction of their targets. It is however a stand-alone tool and can be installed on Unix systems only. The last of such resources, the **UEA sRNA workbench** [17], provides instead a set of originally developed tools, able to address various recurrent tasks (novel ncRNA identification, differential expression analysis, target prediction, etc.) in small ncRNA data analysis. Furthermore, it offers several visualization options, such as secondary structure plots and annotation and alignments display tools.

2.3 Binding Sites Identification

Identifying the targets of an RBP, and the related binding sites is currently performed by RIP-seq (RNA-immunoprecipitation coupled with NGS), HITS-CLIP, PAR-CLIP, and iCLIP (collectively referred to as CLIPs, i.e., cross-linking and immunoprecipitation) or by CRAC/CLASH. While RIP is used to identify RNA targets, the CLIPs can also provide precise binding sites localization on these, eventually allowing the definition of the studied RBP binding specificity (*see* Subheading 4). As these techniques have appeared only in the last few years, analysis methods are still evolving and increasing in terms of variety of approaches. The first two tools we describe, **PIRANHA** [18] and **ASPeak** [19], share the principle of exploiting a coupled expression dataset to help in true binding sites identification. The former, applicable to RIP-seq and all CLIPs, expands this concept to allow any covariate other than expression data, such as genome mappability information or relevant sites (e.g., splice sites) position; it then applies a zero-truncated negative binomial (NB) regression to extract binding sites from the dataset and statistically score them. The latter instead requires a coupled expression dataset, used to compute an expression-sensitive background for binding data; it then runs a NB test over each nucleotide to produce a precise site definition; furthermore, it can exploits the presence of multiple processors to speed up the computation. A last tool able to deal with RIP-seq data is **RIPSeeker** [20], which can also handle PAR-CLIP: its approach is based on stratifying the genome in bins of equal size and applying a two-state Hidden

Markov Model (HMM) with NB emission and a Viterbi algorithm yielding the peak calls, eventually tested for statistical significance. **PIPE-clip** [21] is a Galaxy-based [22] pipeline (and the only web-based tool for binding data analysis) to analyze CLIP experiments: it exploits a zero-truncated NB likelihoods model to identify enriched clusters, then selecting interesting ones by exploiting the assay properties (e.g., mutations for PAR-CLIP), provides preprocessing tools, many customizable parameters and functional annotation of candidate binding sites. The **hyb** [23] pipeline is specially intended for the analysis of CRAC/CLASH data, including tools for preprocessing and mapping reads, and offering features for sensitive CLASH chimeric reads detection and folding; furthermore, this pipeline can also be used for the analysis of CLIP datasets. The last two tools we describe, **PARalyzer** [24] and **wavClusteR** [25], are specifically designed to deal with PAR-CLIP datasets. The former generates two smoothened kernel density estimates, one for T > C transitions and one for non-transitions. Nucleotides with a minimum read depth and a conversion likelihood higher than non-conversion one are considered interaction sites, which are then extended to define the full binding site according to the RBP cross-linking properties; a motif search can also be performed to define the RBP binding motif. The latter tool is instead based on a mixture model, defined on the observed T>C substitutions and the read coverage of the related nucleotides; a continuous wavelet transform is then applied to the model, thus exploiting the coverage function geometry to detect significant discontinuities in coverage, ideally representing true binding sites boundaries.

3 PTR Databases

We now proceed in our exploration of PTR resources by describing the many databases storing and presenting the current knowledge about this class of processes. These tools, listed in Table 2, can be broadly classified according to the kind of data they hold, namely binding sites for *RBP* or *miRNA*, location of *cis-elements* or *ncRNAs* identity and features in general; a few of these resources can be termed *integrative,* in the sense that they present multiple types of information in the same setting, thus allowing for the combination of several PTR facets.

3.1 RBP Binding Sites

These resources collect and present either binding sites for RBPs on mRNAs, derived both by low- and high-throughput experimental approaches, possibly in multiple organisms, or focus on RBP binding specificities, presenting binding motifs and related information. **CLIPZ** [26] is a database and analysis environment for CLIP datasets. Aside from visualizing the included datasets (amounting to ~100 CLIPs plus replicates) at the genome level,

Table 2

Databases presenting available PTR data

Name	Data types	Species	URL	Ref. No.
AURA 2	RBPs, miRNAs, *cis-elements*	Human and mouse	aura.science.unitn.it	[28]
CISBP-RNA	RBPs	Human, mouse, and 22 more	cisbp-rna.ccbr.utoronto.ca	[27]
CLIPZ	RBPs	Human, mouse, and *C. elegans*	clipz.unibas.ch	[26]
doRiNa	RBPs, miRNAs	Human, mouse, *D. melanogaster*, and *C. elegans*	dorina.mdc-berlin.de	[29]
UTRdb/UTRsite	RBPs, miRNAs, cis-elements	Human, mouse, and 77 more	utrdb.ba.itb.cnr.it /utrsite.ba.itb.cnr.it	[31]
lncRNAdb	lncRNAs	Human, mouse, and 58 more	lncrnadb.org	[37]
miRConnX	miRNAs	Human and mouse	benoslab.pitt.edu/mirconnx	[36]
miRecords	miRNAs	Human, mouse, and 7 more	mirecords.biolead.org	[32]
miRGator	miRNAs	Human	mirgator.kobic.re.kr	[34]
miRTarBase	miRNAs	Human, mouse, and 12 more	mirtarbase.mbc.nctu.edu.tw	[33]
miRTCat	miRNAs	Human and mouse	ion.skku.edu/mirtcat	[35]
NONCODE	ncRNAs	Any	noncode.org	[38]
NRED	ncRNAs	Human and mouse	nred.matticklab.com	[39]

RAID	ncRNAs, RBPs	Human	rna-society.org/raid	[40]
Rfam	ncRNAs, cis-elements	Any	rfam.sanger.ac.uk	[41]
starBase 2	ncRNAs, RBPs	Human, mouse, and 4 more	starbase.sysu.edu.cn	[30]
APADB	Alternative polyadenylation sites	Human, mouse, and chicken	tools.genxpro.net/apadb	[42]
ARED	AU-rich elements	Human and mouse	rc.kfshrc.edu.sa/ared	[43]
AREsite	AU-rich elements	Human and mouse	rna.tbi.univie.ac.at/AREsite	[44]
IRESite	Internal ribosome entry sites	Human, mouse, and 7 more	iresite.org	[45]
SelenoDB	SECIS elements	Human, mouse, and 6 more	selenodb.org	[46]
SIREs	Iron-responsive elements	Any	ccbg.imppc.org/sires	[47]
Transterm	*cis*-elements	Any	mrna.otago.ac.nz	[48]

The table lists databases containing data about PTR such as RBP and miRNA binding sites and *cis*-elements regions. Shown are the data types (i.e., RBP and noncoding RNA binding sites, all *cis*-elements or a specific type of these) provided by each database, the species for which data is available and how these databases can be reached (URL and reference)

users can upload and analyze their own data down to the identification of enriched binding motifs and binding sites on the mRNAs; furthermore, miRNA-specific tools are provided to handle Argonaute CLIPs used to identify targets for these ncRNAs. **CISBP-RNA** [27] focuses instead on RBP motifs, collecting many experiments either from the literature or performed by the laboratory of the database maintainers. As multiple techniques and organisms are included, one can obtain a precise picture of the binding specificities for the RBP of interest; furthermore, tools are provided to predict instances of an RBP binding motif in RNA sequences or to compare custom motifs with the ones stored in the database.

The other resources presenting RBP binding sites, namely **AURA2** [28], **doRiNa** [29], **starBase2** [30], and **UTRdb/UTRsite** [31], also provide other types of data (e.g., miRNA binding sites) and are thus described in the integrative resources section (Subheading 3.4).

3.2 miRNA Binding Sites and ncRNAs

These databases aim at presenting data about noncoding RNAs such as their identity, role, and molecular targets. The majority of these is focused in particular on collecting miRNA–mRNA interactions, both derived experimentally and computationally through the plethora of available predictors; the functions of individual miRNAs are also often presented, derived by analyzing processes and pathways in which their targets are involved.

miRecords [32] and **miRTarBase** [33] provide a curated collection of several thousand experimentally verified and/or computationally predicted miRNA–mRNA interactions (only in miRecords, compiled by 11 different algorithms); miRTarBase also provides miRNA and targets expression profiles, data about miRNA association to diseases and the possibility to trace miRNA–targets networks. On the same line but heavily focused on NGS data, **miRGator** [34] also provide a catalog of NGS-derived miRNAs for various tissues and organs, a dedicated browser concurrently displaying miRNA sequencing data and secondary structure and miRNA–target expression correlations, all combined with the goal of identifying true regulatory, functional, and pathology-related associations. **miRTCat** [35] provides miRNA binding sites derived by AGO HITS-CLIP along with prediction of binding sites in 3′ UTR regions (based on sequence conservation): its peculiarity lies in including a novel type of noncanonical target sites type, the *non-nucleation bulge*, which was first discovered in AGO HITS-CLIP data. The specificity of **miRConnX** [36] is instead that it allows users to upload their expression data: these are then matched with precomputed TF–gene and miRNA–gene genome-wide networks (derived by predictions and supplemented with known interactions), to ultimately derive a condition-specific miRNA–mRNA regulatory network, useful for hypothesis generation and further analyses.

Among resources dedicated to ncRNAs features or to a specific class of these, **lncRNAdb** [37] and **NONCODE** [38] are dedicated to long noncoding RNAs (lncRNAs) including annotations such as sequence and structure information, expression profiles, conservation and function for multiple organisms. Both databases also include links to external resources, such as **NRED** [39], allowing to retrieve further information about the functions and the expression of these RNAs.

RAID [40] focuses on providing a literature-curated catalog of RNA-associated interactions, including both RNA–RNA and RNA–protein interactions. This data potentially help understanding the role, based on their binding properties, of various ncRNA molecules still not fully characterized. The interactions are classified by molecule type of the participants (e.g., lncRNA-associated, or snoRNA-associated), and annotations such as binding site, tissue type, experimental technique, and others are included.

Eventually, **Rfam** [41] presents a very wide set of RNA families describing the various RNA gene types (including miRNAs, lncRNAs, snoRNAs, and many others) and mRNA cis-elements. Families are defined through covariate models by sequence alignments and primarily having a conserved structure. Entries are extensively described in a Wikipedia-like format with description, figures and references; furthermore, a tool is provided to allow the user to scan RNA sequences, thus identifying matches with Rfam families.

The other resources presenting miRNA binding sites or ncRNAs data, namely **AURA2** [28], **doRiNa** [29], **starBase2** [30], and **UTRdb/UTRsite** [31], also provide other types of data (e.g., RBP sites or cis-elements) and thus are described in the integrative resources section (Subheading 3.4).

3.3 Cis-Elements

These resources usually focus on one or multiple types of cis-elements and aim at presenting related features such as their instances on mRNAs, factors binding to and mediating the role of the element; several databases also include predictive tools to help in identifying previously undetected instances of these elements.

APADB [42] provides 3′end sequencing-derived information about alternative polyadenylation sites in 3′ UTRs, including both coding and noncoding genes. The data is displayed through a genome browser, organized by tissue/organ and available for human and other two organisms; furthermore, potential losses of miRNA binding sites are also highlighted to help in understanding regulatory changes due to alternative polyadenylation events. **ARED** [43] and **AREsite** [44] are two databases focusing on AU-rich Elements (ARE), a well-characterized cis-element commonly found in 3′ UTRs of mRNAs and involved in their stability (through the action of several RBPs termed ARE-binding proteins, or ARE-BPs). Both resources provide a computationally mapped catalog of AREs, obtained by matching one or more sequence

patterns (one for ARED and eight for ARESite). Both databases offer some degree of annotation with ARESite, the most complete, including a graphical representation of found AREs, structural information, phylogenetic conservation and supporting evidence extracted from the literature. **IRESite** [45] aims instead at producing a curated catalog of all experimentally known internal ribosome entry sites (IRES), both cellular and viral: these elements mediate translation initiation without the need of a 5′cap structure, thus allowing protein synthesis in stress conditions and of viral mRNAs. The database provides detailed information about each IRES (sequences, translation efficiency, conditions, etc.), extracted from the literature, along with tools to search custom sequences against known IRESs to detect potentially novel instances of this element.

As the next two resources underscore, mRNA cis-elements are also considerably involved in the metabolism of several chemical elements. **SelenoDB** [46] is indeed a resource devoted to the description of selenogenes and selenocysteine insertion sequence (SECIS) elements. This element, found in the 3′ UTRs, recruits proteins involved in selenium metabolism to the mRNA through its characteristic stem-loop structure. Instances of this element were computationally predicted and then manually curated: these are graphically displayed and correlated with several annotations. **SIREs** [47] is instead a web server for the prediction of iron-responsive elements (IREs), specific cis-elements found in the mRNA of proteins involved in iron metabolism; this element is well characterized in both its sequence and secondary structure. This resource allows users to input their own sequence and, based on patterns derived from this characterization, will predict IRE positions, features, and specificity for one of the iron responsive proteins (IRP1 or IRP2) binding to these elements.

Transterm [48] is a database of regions affecting translation, including both experimentally validated regulatory elements and the ability to scan a user-inputted sequence to identify instances of the many cis-elements classes for which a searchable pattern could be defined (extracted from the literature). A detailed description of the various elements classes is provided, as are several basic annotations on the genomes which can be analyzed on the website (e.g., initiation codon context).

The other resources presenting data about cis-elements, namely **AURA2** and **UTRdb/UTRsite** [31], also provide other types of data (e.g., RBP binding sites) and thus are described in the integrative resources section (Subheading 3.4).

3.4 Integrative

We define a PTR database as *integrative* if it collects data about multiple aspects of regulation such as for instance RBP- and miRNA-mediated regulation: the principle behind these resources is to more precisely and completely define an mRNA potential for

PTR, through the parallel observation of many factors possibly mediating this potential: such an approach would eventually allow users to make the most of publicly available data. However, collecting such amounts and variety of data is a daunting task, often requiring manual literature searching; indeed, just a few resources following this approach are currently available.

AURA2 [28] is a meta-database focused on the UTRs and providing data regarding the multiple aspects of PTR these regions mediate. It includes experimentally derived RBP and miRNA binding sites, cis-elements, RNA methylation and SNP data, complemented with multiple annotations such as phylogenetic conservation, secondary structures and functional descriptions. A custom UTR browser, along with several additional views and batch tools, allows the users to display all the various datasets at the same time, thus helping in obtaining a complete understanding of the regulation to which a UTR is subjected. Also **UTRdb/UTRsite** [31] focus on 5′ and 3′ UTRs, providing annotations for these regions in many different organisms. It includes instances of cis-elements (including polyadenylation signals), phylogenetic conservation, SNPs, and experimentally determined miRNA targets; RBP binding sites are however absent from this resource. Furthermore, the UTRsite section works on custom sequences predicting the occurrence of many cis-element types.

doRiNa [29] is a database dedicated to RNA interactions, including both RBP and miRNA binding sites as derived by CLIP approaches; miRNA targets are also provided in a computationally predicted form (by means of several algorithms); searches can be performed by selecting a trans-factor and an mRNA region of interest: detailed results can then be displayed with the help of a genome browser. Furthermore, the combinatorial search tool allows the identification of mRNAs regulated by multiple factors of interest. Also **starBase2** [30] is a database exploiting exclusively CLIP data; however it aims at identifying interactions between miRNAs and several other types of RNA, namely mRNA, lncRNA, ceRNA, circRNA, and other noncoding RNAs; it also includes protein–RNA interactions derived from the same sources. Interactions are displayed by category and annotated with expression data; furthermore, miRNA and ceRNA functions can be predicted through dedicated tools leveraging on functional ontologies terms.

4 Prediction Tools

When PTR data about genes of interest are not available, or were produced in too much different systems and conditions to be integrated, prediction tools can come to help in formulating a biological hypothesis. While some problems are relatively easy to address, and thus a lot of tools are available (e.g., miRNA target

identification), others are particularly complicated (e.g., RBP binding sites) and just a few dedicated tools have emerged up to now. We now describe the various methods allowing to predict the presence of regulatory elements and their role in PTR: these tools are listed in Table 3 and can be broadly grouped by the type of prediction they provide, namely identifying *RBP targets*, *miRNA targets* or the effects of *SNPs on RNA secondary structures*.

4.1 RBP Targets

Predicting the targets of an RBP and the location of its binding sites on RNA molecules is known to be a challenging task. Both identifying which residues of a protein may bind RNA and which sequence or structure specificities these residues may confer are complex problems for which no precisely defined rules exist. Indeed, only a few tools have tried addressing this problem so far.

catRapid [49] is a webserver offering an algorithm performing de novo prediction of protein–RNA interactions based on physico-chemical properties of polypeptides and nucleotide chains. The interaction propensity of these molecules is thus calculated solely based on their sequence: in particular, secondary structure, hydrogen bonding and van der Waals propensities are computed and combined together to yield an interaction profile; eventually an interaction propensity score and an evaluation of the interaction statistical significance derived by a discriminative power calculation are returned. **RBPmap** [50] is instead a tool based on a weighted-rank approach, accepting any RNA sequence as input and exploiting currently available RBP binding motifs, extracted from the literature. It is optimized for human, mouse and fruit fly, although other organisms are supported too. The algorithm matches the motifs matrices to the user input sequence; it then takes into account the propensity of binding sites for clustering and the overall conservation of the region in order to guide the identification of true binding sites. The last resource belonging to this category, **ScanForMotifs** [51], is a webserver enabling the prediction of RBP binding sites, miRNA targets and cis-elements by means of a set of 3′ UTR alignments, known RBP binding motifs, miRNA seeds, and Transterm [48] elements. Users can provide a gene symbol or a sequence alignment and tune a few statistical parameters: the tool will then run three parallel jobs to deal with each prediction type; results will eventually graphically show each identified site on the input alignment.

4.2 miRNA Targets

Contrary to predicting an RBP binding sites, identifying targets for a miRNA may seem to be a quite straightforward task: once the miRNA seed sequence is known, the problem consists in finding matching complementary sequences in the mRNAs. However, this intuitive procedure has been proven to produce an high number of false positives and also to miss noncanonical miRNA binding sites. Nevertheless, the accessible nature of this task has led to the

Table 3

Prediction tools for posttranscriptional regulation events

Name	Prediction type	Approach	URL	Ref. No.
catRapid	RBP targets	Computes interaction propensity distribution and ranks the results	service.tartaglialab.com/page/catrapid_group	[49]
RBPmap	RBP targets	Uses a weighted-rank algorithm considering sites clustering propensity and conservation	rbpmap.technion.ac.il	[50]
ScanForMotifs	RBP and miRNA targets, cis-elements	Combines multiple data sources to predict regulatory elements in 3′ UTRs	bioanalysis.otago.ac.nz/sfm	[51]
ComiR	combinatorial miRNA targeting	Uses thermodynamic modeling and machine learning techniques coupled to expression data	benoslab.pitt.edu/comir	[57]
DIANA-microT	miRNA targets	Predicts targets by exploiting positive and negative recognition element sets as defined by PAR-CLIP data	microrna.gr/webServer	[54]
miRanda	miRNA targets	Employs sequence position-specific and conservation rules	microrna.org	[59]
MAGIA2	miRNA–TF regulatory networks	Integrates targets prediction and expression data	gencomp.bio.unipd.it/magia2	[58]
miRmap	miRNA targets	Ranks by repression strength through the use of multiple features	mirmap.ezlab.org	[53]
PicTar	miRNA targets	Employs vertebrates alignment to identify targets	pictar.mdc-berlin.de	[60]
PITA	miRNA targets	Exploits energy changes in miRNA–target duplex formation to predict targeting	genie.weizmann.ac.il/pubs/mir07/	[61]
TargetProfiler	miRNA targets	Uses an HMM trained on experimentally verified miRNA targets	mirna.imbb.forth.gr/Targetprofiler.html	[56]
TargetScan	miRNA targets	Predict targets by identifying conserved seed complementarity	targetscan.org	[52]
RNAsnp	SNP effects on RNA structure	Computes SNP effect on ensemble of structures and deriving p-values	rth.dk/resources/rnasnp/	[62]
SNPfold	SNP effects on RNA structure	Use a partition function calculation that considers the ensemble of possible RNA conformations	ribosnitch.bio.unc.edu/snpfold	[63]

The table lists tools for prediction of posttranscriptional phenomena such as RBP and miRNA binding and SNP effect on RNA secondary structure. Listed are the prediction type, the approach adopted by the tool and how to retrieve it (URL and reference)

development of many tools, exploiting different principles to attempt at discriminating true sites from the bulk.

TargetScan [52] is an algorithm predicting miRNA targets by exploiting phylogenetic conservation information over 46 vertebrate species: under the assumption that conservation often implies function, conserved sites (7-mer and 8-mer) that match the seed region of miRNAs are extracted and used to associate a miRNA to its target mRNAs. Predictions are then ranked by a targeting efficacy score, determined by keeping into account various features of the site context. Going beyond seed match identification alone, **miRmap** [53] is a webserver based on a Python library employing thermodynamic, evolutionary and sequence-based features. These features are then combined by means of a linear model to eventually yield the '*miRmap score*', representing the predicted miRNA repression strength. The library can be integrated in other applications through a REST service (i.e., allowing to access its data programmatically by composing specific URLs), or the precomputed predictions be downloaded in full from the website. **DIANA-microT** [54] is a webserver detecting miRNA sites both in 3′ UTRs and in coding sequences: the algorithm is trained on a positive and negative miRNA-recognition element set defined by an Argonaute PAR-CLIP assay. Potential 3′ UTR and coding sequence sites are considered separately, and specific features (including conservation, flanking AU content, and others) are computed for every candidate; these are then combined and eventually scored by generalized linear models. Furthermore, this resource offers a useful plugin for the Taverna [55] workflow platform, allowing the inclusion of miRNA target prediction into an analysis pipeline. Also **TargetProfiler** [56] exploits a small set of experimentally derived miRNA targets to train a model, a Hidden Markov Model (HMM) in this case, then used to probabilistically learn miRNA–target associations. Predicted targets are then filtered according to the HMM score, the miRNA–mRNA hybrid free energy and the eight-species phylogenetic conservation of the site region. The webserver provides precomputed predictions for all human genes and miRNAs.

Rather than focusing on a single miRNA, **ComiR** [57] aims at computing the potential of mRNAs to be regulated by a set of miRNAs, directly provided by the user or derived by input expression levels. By employing four different methods, complementing each other and integrating expression levels, the tool first computes miRNA–mRNA interaction probabilities for each miRNA, then integrating these probabilities by an SVM model; the output is a list of genes, ranked by the probability of being targets of the miRNA set. Eventually, **MAGIA2** [58] is a tool integrating miRNA target prediction with miRNA and gene expression to attempt at reconstructing transcription factors and miRNA-mediated regulatory networks from the input data. A selectable

correlation measure is computed (TF–gene and miRNA–gene) and integrated with multiple regulatory predictions, exploiting many tools for miRNA target predictions and two databases of miRNA–TF and TF–gene associations. So derived regulatory circuits are then dynamically displayed through tables and graphical network views.

A few other tools devoted to miRNA targets prediction, namely **miRanda** [59], **PicTar** [60], and **PITA** [61], are not described in detail here as they were no longer updated in the last 3 years: these are nevertheless listed in Table 3 along with all other tools.

4.3 SNPs Impact on RNA Secondary Structure

The effects of genetic variation are most often studied on protein-coding sequences only (e.g., exome sequencing), thus focusing on changes in protein domains and related features. However, variants in the noncoding portions of an mRNA may heavily affect the regulation of these transcripts, thus altering the abundance of an otherwise functional protein. To study the impact of such variants a couple of tools are now available, investigating structural consequences connected to the presence of SNPs.

The first tool, **RNAsnp** [62], is a web server based on computing the structural differences between wild-type and mutated sequence by means of an RNA folding method and a dynamic programming algorithm over windows of fixed length. The tool includes three algorithms; one is tuned for short sequences (<1000nt), one for long ones and a last method consisting in the combination of the other two approaches; all three algorithms report as output the window of maximum base pair distance and the related p-value. The other available tool, **SNPfold** [63], first computes a partition function (i.e., a matrix representing the probability of base pairing for all possible pairs in the sequence) over the wild-type and the mutated sequence; then, it determines how much the two structure differs (by means of a correlation coefficient) and also where this difference is the greatest. By means of this partition function, the ensemble of all possible structural conformations for both wild-type and mutated sequences are considered, thus reporting results with more confidence and a more reliable *p*-value.

5 Tools for RNA Sequence and Structure Motif Search

We conclude our tool presentation by describing a set of resources, listed in Table 4, which are aimed at *sequence* and/or *structural* RNA motifs identification: this type of task is particularly frequent and is needed to, for instance, determine binding preferences for an RBP or identify what regulatory element may be responsible for a shared translational control pattern observed in a group of genes

Table 4

Tools for RNA motif search

Name	Motif type	Approach	URL	Ref. No.
CMfinder	Sequence	Expectation maximization algorithm using covariate models to describe motifs	bio.cs.washington.edu/yzizhen/CMfinder	[64]
MEMERIS	Sequence/structure	Expectation maximization algorithm which uses secondary structure to guide search towards single-stranded RNA regions	bioinf.uni-freiburg.de/~hiller/MEMERIS	[65]
RBPmotif	Sequence/structure	Uses binding data to predict RBP binding specificities and sequence motifs to predict structural preferences	rnamotif.org	[69]
RNAcontext	Sequence/structure	Computes structural context distributions on bound sequences to train a model	morrislab.med.utoronto.ca/software	[70]
RNAmotifs	Multivalent RNA motifs	Identifies multivalent RNA motifs from analysis of differentially regulated exons	bitbucket.org/rogrro/rna_motifs	[71]
RNApromo	Structure	Uses a stochastic context-free grammar and a probabilistic inference algorithm	genie.weizmann.ac.il/pubs/rnamotifs08/rnamotifs08_predict.html	[67]
TEISER	Structure	Finds enriched/depleted structures that explain patterns in genome-wide data	tavazoielab.c2b2.columbia.edu/TEISER	[68]

The table lists algorithms for RNA motifs identification, both at the sequence and at the secondary structure levels. Searched motif type, the adopted approach and how to retrieve each tool (URL and reference) are indicated

(e.g., as identified by polysome profiling). While the majority of tools outputs motifs focused on one of the two aspects (either sequence or structure), most algorithms take a step further by adopting an integrative approach which considers both aspects to define the motifs.

CMfinder [64] is a tool based on an expectation-maximization (EM) algorithm, including RNA secondary structure information by using covariance models (CM). While the EM algorithm drives the search, motifs distribution in sequences is described by a mixture model, and motifs themselves are modeled by a CM; given the complexity of the so-defined search space, the algorithm uses heuristics to select interesting candidates. In particular only motifs with stable secondary structures are considered and then aligned to define the motif consensus. Results are eventually refined by a second EM algorithm to yield the final predicted motifs. Exploiting EM algorithms as well, **MEMERIS** [65] is a tool based on the popular MEME motif search software [66], which identifies sequence motifs by guiding the search towards single-stranded regions; this criterion is justified by the preference of several RBPs for binding to such regions. The guidance is made possible by replacing the MEME uniform motif start probability distribution by a single-strandedness distribution computed on the input RNA sequence; furthermore, in order to allow maximal flexibility, the weight of the single-strandedness assumption can be tuned to the user taste. **RNApromo** [67] is an algorithm instead based on stochastic context-free grammars (SCFGs), devoted to the identification of short secondary structure motifs in RNA sequences: to reduce the search space, the algorithm requires a suggested structure as input, along with the set of sequences supposingly sharing such motif. The algorithm first identifies a set of structures that appear in as many sequences as possible; these are then refined and statistically evaluated by means of a probabilistic inference algorithm. Also based on context-free grammars, **TEISER** [68] is a framework aimed at discovering structural motifs that can be correlated with genome-wide measurements such as, to cite one, mRNA stability data: the tool use mutual information to understand the impact of presence/absence of many possible structural elements on the provided measurements. It is thus possible, for instance, to deduce the dependency of mRNA stability on the presence of a specific hairpin in the mRNA 5′ or 3′ UTRs. Candidate motifs are then refined by selecting the ones with the greatest impact on such measurements, which are eventually statistically assessed to yield truly relevant motifs.

The last tools we describe are devoted to a specific motif identification problem, rather than being generally applicable to any set of RNA sequences. **RBPmotif** [69] is a webserver focused on discovering the sequence and structural binding preferences of RBPs. If no such preference is known, the tool will run an algorithm

(**RNAcontext** [70], requiring a set of bound and unbound sequences as input) to investigate on this aspects; on the other end, if a sequence motif is already available for the RBP, an additional analysis will be run to identify potential structural contributions to the RBP sequence binding preference. Statistically evaluated motifs are eventually returned and can be compared either by considering motif instances in bound and unbound sequences or by looking at similar binding motifs of other RBPs. Eventually, **RNAmotifs** [71] stands out of the pack because of its particular application: indeed, this tool is aimed at identifying motifs involved in splicing regulation of a set of differentially regulated alternative exons. In particular, motifs are defined as either degenerate or nondegenerate tetramers found around enhanced or repressed exons: these tetramers are tested for enrichment in sequences surrounding these exons, and statistically evaluated by a Fisher test and a bootstrap procedure. Furthermore, splicing maps derived by the enriched tetramers score profiles can also be visualized.

6 Pipelines for PTR: Two Case Studies

The many available tools and databases we described, which can be used to analyze various types of PTR data, all are individually useful and serve a purpose on their own relating to a specific analysis task. Nevertheless, to get the most out of the data deluge coming out from such genome-wide approaches and reach the highest resolution and accuracy level that the resulting datasets enable, these tools must be combined and integrated into full-blown analysis workflows. Toward this goal, we thus now conclude the chapter by presenting two tentative pipelines, combining several of these resources to address the analytic needs of two different usage cases. Through these examples, illustrated in Fig. 1, this section aims at providing initial practical guidance to researchers approaching for the first time PTR data analysis.

6.1 Case Study 1: Impact of a Treatment on Translation

Our first case study deals with a particularly common experiment in PTR, consisting in profiling the effects of treatment/stimulus on translational control in a suitable cellular model system: the ultimate goal is understanding which regulatory factors/elements influence translatability of the transcripts following the treatment. This task is often addressed by means of coupled total and polysomal RNA extraction followed by an RNA-seq assay on the poly(A+) fraction, eventually allowing quantification of translational efficiency for each individual transcript. In such an experimental setting, the analysis can be subdivided in five phases, illustrated by Fig. 1a. First of all, one needs to identify genes (called DEGs for differentially expressed genes) which are significantly changing their translation levels following the treatment: this task

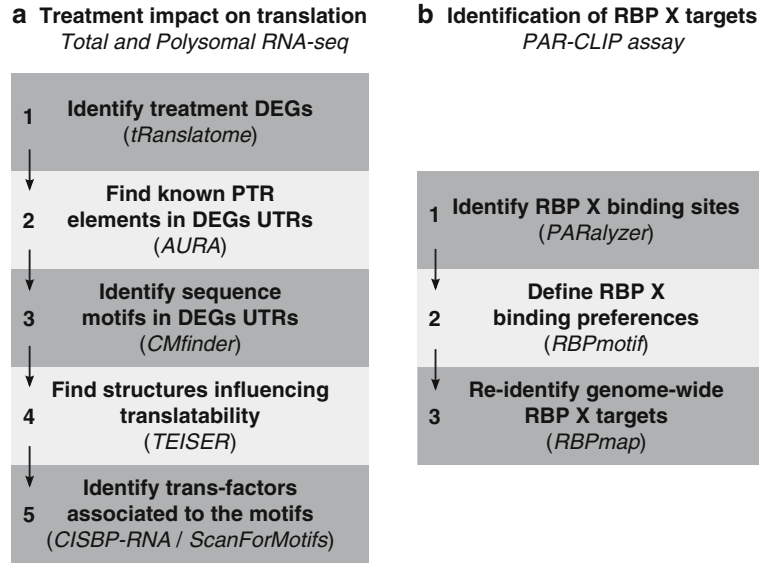

Fig. 1 Representative data analysis pipelines to address two PTR case studies. The figure displays two potential data analysis pipeline for two common types of experiment in post-transcriptional regulation of gene expression. (**a**) Describes a potential pipeline for the analysis of a study to determine the impact of a treatment/stimuli on translation. This can be performed by total and polysomal profiling coupled with RNA-seq. Differentially translated genes (DEGs) can first of all be identified by **tRanslatome**; known PTR elements influencing translation (RBP binding sites, cis-elements, etc.) can then be identified by **AURA2** in DEGs UTRs. Next, two motif search analyses can be performed to identify potential determinants of the observed translational changes, at the sequence level (**CMfinder**) and structure-wise (**TEISER**). Eventually, these motifs may be characterized by attempting at identifying trans-factors binding to them by means of **CISBP-RNA** and/or **ScanForMotifs**. (**b**) describes a workflow aimed at the identification of targets and binding preferences for a generic RBP, named X. Starting from a PAR-CLIP assay in the system of interest, binding sites for the RBP are first identified by means of **PARalyzer**; subsequently, its sequence and structure binding specificities are determined through the **RBPmotif** tool. Eventually, the targets list is reassessed by predicting binding sites (using the motif computed in the previous step) with **RBPmap**, in order to further confirm PAR-CLIP results and identify additional targets which may be non-expressed in the studied system (and thus potentially missed by the assay)

can be performed by **tRanslatome** [13], by providing a table with per-gene read counts (steps required to obtain these counts from raw reads are related to the RNA-seq technique and are thus not described here; readers can refer to this review [72]) as input and choosing one of the available methods for DEGs calling. The output will consist of a list of genes with significant changes, representing the treatment-induced biology. The next step aims at obtaining a first overview of which trans-factor binding sites and

cis-elements are already known to be present in the mRNAs UTRs of DEGs: this task can be performed by inputting the gene list to the batch tools (here the *Regulatory Element Enrichment* one) of **AURA2** [28]; these results could also be complemented by predicting the presence of such elements through a tool such as **ScanForMotifs** [51]. At the end of this step, one should already be able to understand whether one or a few trans-factors/cis-elements are massively shared by DEGs, and thus, if this is the case, decide to focus on these as interesting candidates to explain the treatment impact. If this is not the case, the next step consists in looking for shared sequence motifs in DEG UTRs: this analysis can be efficiently performed by providing DEGs 5′ and 3′ UTRs sequences (in two distinct executions) to **CMfinder** [64]; a further step in the same direction, which can be addressed by means of **TEISER** [68], consists in detecting structural elements which can explain DEG changes in translational efficiency (to do so, one can provide a list of genes and their related translational fold-changes to the tool; sequences are not needed, at least for the most common organisms, as this tool comes with such data for a number of prepackaged genomes). The results of these two steps will eventually yield sequence and structure motifs shared by a consistent number of DEG UTRs and which represent the distinctive post-transcriptional candidates acted upon to obtain the treatment effects. Thus, as a last step, these motifs should be analyzed to understand which trans-factors are targeting them to mediate these effects: to do so, one can try to find matches with known binding motifs or cis-elements patterns by means of **CISBP-RNA** [27] and **ScanForMotifs** [51]; obviously, these analyses should be coupled to a thorough literature search, especially for structure motifs, to maximize the chances of identifying the factors at play.

6.2 Case Study 2: Identification of an RBP Targets and Binding Preferences

Our second case study focuses instead on an increasingly common kind of PTR experiment, namely identifying targets and binding preferences for an RBP. This kind of experiment has recently been made possible by the advent of the CLIP family of techniques. In particular, our case study focuses on the PAR-CLIP technique, exploiting T>C conversion induced by retrotranscription of the incorporated photoactivatable 4SU nucleosides to precisely pinpoint the binding sites. The analysis workflow for such an experiment, illustrated in Fig. 1b, can be subdivided in three phases. The first phase consists in processing aligned reads to detect "peaks" (i.e., mRNA regions hosting a significantly greater number of reads with respect to the genome-wide background) and intersect these with T>C conversion sites. This processing, done by means of **PARalyzer** [24], produces a list of binding sites for our RBP of interest, along with the estimation of their statistical significance. From the list the set of targets for this RBP can be implicitly inferred, along with processes and pathways in which these are

involved. We can thus obtain an overall view of the potential role that this RBP has in the studied system. While this information is extremely useful, we still lack, at this point, a definition of how this RBP chooses the regions it targets on the mRNAs. To address this issue, the next step, performed by means of the **RBPmotif** [69] tool, exploits the set of bound sequences (and a corresponding set of unbound ones, also derived from the PAR-CLIP assay) to learn the binding preference of the RBP and derive both a sequence logo and a structural preference indicator. That done, we are still left with a last potential issue: our assay was performed in a single model system, expressing a specific set of genes, considerably smaller than the total number of protein-coding loci in the genome: we are thus missing a fraction of all potential targets, due to them not being expressed in that system. To alleviate this issue, we could exploit the binding preferences we have derived to predict, over the whole transcriptome, potential binding sites for our RBP: this task can be performed by means of **RBPmap** [50], providing the sequence motif (either its consensus sequence or its probability matrix) as input to the tool. This analysis will also allow us to compute the goodness of fit between experimentally derived binding sites and the predicted motif matches, thus eventually enabling the evaluation of the motif quality and possibly its refinement.

References

1. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31–46
2. Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. Nat Rev Genet 15:205–213
3. Milek M, Wyler E, Landthaler M (2012) Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin Cell Dev Biol 23:206–212
4. Arava Y (2003) Isolation of polysomal RNA for microarray analysis. Methods Mol Biol (Clifton, NJ) 224:79–87
5. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215
6. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
7. Konig J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–915
8. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665
9. Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. Nat Biotechnol 27:667–670
10. Lambert N, Robertson A, Jangi M, McGeary S, Sharp PA, Burge CB (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Mol Cell 54:887–900
11. Campbell ZT, Bhimsaria D, Valley CT, Rodriguez-Martinez JA, Menichelli E, Williamson JR, Ansari AZ, Wickens M (2012) Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. Cell Rep 1:570–581
12. Larsson O, Sonenberg N, Nadon R (2011) anota: Analysis of differential translation in genome-wide studies. Bioinformatics (Oxford, England) 27:1440–1441

13. Tebaldi T, Dassi E, Kostoska G, Viero G, Quattrone A (2014) tRanslatome: an R/Bioconductor package to portray translational control. Bioinformatics (Oxford, England) 30:289–291
14. Olshen AB, Hsieh AC, Stumpf CR, Olshen RA, Ruggero D, Taylor BS (2013) Assessing gene-level translational control from ribosome profiling. Bioinformatics (Oxford, England) 29:2995–3002
15. Zhang Y, Xu B, Yang Y, Ban R, Zhang H, Jiang X, Cooke HJ, Xue Y, Shi Q (2012) CPSS: a computational platform for the analysis of small RNA deep sequencing data. Bioinformatics (Oxford, England) 28:1925–1927
16. Giurato G, De Filippo MR, Rinaldi A, Hashim A, Nassa G, Ravo M, Rizzo F, Tarallo R, Weisz A (2013) iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. BMC Bioinformatics 14:362
17. Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L, Schwach F, Dalmay T, Moulton V (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics (Oxford, England) 28: 2059–2061
18. Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD (2012) Site identification in high-throughput RNA-protein interaction data. Bioinformatics (Oxford, England) 28:3013–3020
19. Kucukural A, Ozadam H, Singh G, Moore MJ, Cenik C (2013) ASPeak: an abundance sensitive peak detection algorithm for RIP-Seq. Bioinformatics (Oxford, England) 29:2485–2486
20. Li Y, Zhao DY, Greenblatt JF, Zhang Z (2013) RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. Nucleic Acids Res 41, e94
21. Chen B, Yun J, Kim MS, Mendell JT, Xie Y (2014) PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. Genome Biol 15:R18
22. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol/edited by Frederick M Ausubel [et al] Chapter 19, Unit 19 10:11–21
23. Travis AJ, Moody J, Helwak A, Tollervey D, Kudla G (2014) Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. Methods 65:263–273
24. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79
25. Sievers C, Schlumpf T, Sawarkar R, Comoglio F, Paro R (2012) Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. Nucleic Acids Res 40:e160
26. Khorshid M, Rodak C, Zavolan M (2011) CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic Acids Res 39:D245–D252
27. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A, Na H, Irimia M, Matzat LH, Dale RK, Smith SA, Yarosh CA, Kelly SM, Nabet B, Mecenas D, Li W, Laishram RS, Qiao M, Lipshitz HD, Piano F, Corbett AH, Carstens RP, Frey BJ, Anderson RA, Lynch KW, Penalva LO, Lei EP, Fraser AG, Blencowe BJ, Morris QD, Hughes TR (2013) A compendium of RNA-binding motifs for decoding gene regulation. Nature 499:172–177
28. Dassi E, Re A, Leo S, Tebaldi T, Pasini L, Peroni D, Quattrone A (2014) AURA 2: empowering discovery of post-transcriptional networks. Translation 2:e27738
29. Anders G, Mackowiak SD, Jens M, Maaskola J, Kuntzagk A, Rajewsky N, Landthaler M, Dieterich C (2012) doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic Acids Res 40:D180–D186
30. Li JH, Liu S, Zhou H, Qu LH, Yang JH (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 42:D92–D97
31. Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 38:D75–D80
32. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res 37:D105–D110
33. Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, Chu CF, Huang HY, Lin CM, Ho SY, Jian TY, Lin FM, Chang TH, Weng SL, Liao KW, Liao IE, Liu CC, Huang HD (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 42:D78–D85

34. Cho S, Jang I, Jun Y, Yoon S, Ko M, Kwon Y, Choi I, Chang H, Ryu D, Lee B, Kim VN, Kim W, Lee S (2013) MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. Nucleic Acids Res 41:D252–D257
35. Kim KK, Ham J, Chi SW (2013) miRTCat: a comprehensive map of human and mouse microRNA target sites including non-canonical nucleation bulges. Bioinformatics (Oxford, England) 29:1898–1899
36. Huang GT, Athanassiou C, Benos PV (2011) mirConnX: condition-specific mRNA-microRNA network integrator. Nucleic Acids Res 39:W416–W423
37. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS (2011) lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res 39:D146–D151
38. Bu D, Yu K, Sun S, Xie C, Skogerbo G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, Zhao H, Liu Z, Liu C, Chen R, Zhao Y (2012) NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res 40:D210–D215
39. Dinger ME, Pang KC, Mercer TR, Crowe ML, Grimmond SM, Mattick JS (2009) NRED: a database of long noncoding RNA expression. Nucleic Acids Res 37:D122–D126
40. Zhang X, Wu D, Chen L, Li X, Yang J, Fan D, Dong T, Liu M, Tan P, Xu J, Yi Y, Wang Y, Zou H, Hu Y, Fan K, Kang J, Huang Y, Miao Z, Bi M, Jin N, Li K, Li X, Xu J, Wang D (2014) RAID: a comprehensive resource for human RNA-associated (RNA-RNA/RNA-protein) interaction. RNA (New York, NY) 20:989–993
41. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226–D232
42. Muller S, Rycak L, Afonso-Grunz F, Winter P, Zawada AM, Damrath E, Scheider J, Schmah J, Koch I, Kahl G, Rotter B (2014) APADB: a database for alternative polyadenylation and microRNA regulation events. Database 2014:bau076
43. Bakheet T, Williams BR, Khabar KS (2006) ARED 3.0: the large and diverse AU-rich transcriptome. Nucleic Acids Res 34:D111–D114
44. Gruber AR, Fallmann J, Kratochvill F, Kovarik P, Hofacker IL (2011) AREsite: a database for the comprehensive investigation of AU-rich elements. Nucleic Acids Res 39:D66–D69
45. Mokrejs M, Masek T, Vopalensky V, Hlubucek P, Delbos P, Pospisek M (2010) IRESite--a tool for the examination of viral and cellular internal ribosome entry sites. Nucleic Acids Res 38:D131–D136
46. Castellano S, Gladyshev VN, Guigo R, Berry MJ (2008) SelenoDB 1.0: a database of selenoprotein genes, proteins and SECIS elements. Nucleic Acids Res 36:D332–D338
47. Campillos M, Cases I, Hentze MW, Sanchez M (2010) SIREs: searching for iron-responsive elements. Nucleic Acids Res 38:W360–W367
48. Jacobs GH, Chen A, Stevens SG, Stockwell PA, Black MA, Tate WP, Brown CM (2009) Transterm: a database to aid the analysis of regulatory sequences in mRNAs. Nucleic Acids Res 37:D72–D76
49. Agostini F, Zanzoni A, Klus P, Marchese D, Cirillo D, Tartaglia GG (2013) catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinformatics (Oxford, England) 29:2928–2930
50. Paz I, Kosti I, Ares M Jr, Cline M, Mandel-Gutfreund Y (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res 42:W361–W367
51. Biswas A, Brown CM (2014) Scan for Motifs: a webserver for the analysis of post-transcriptional regulatory elements in the 3' untranslated regions (3' UTRs) of mRNAs. BMC Bioinformatics 15:174
52. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20
53. Vejnar CE, Blum M, Zdobnov EM (2013) miRmap web: comprehensive microRNA target prediction online. Nucleic Acids Res 41:W165–W168
54. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res 41:W169–W173
55. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, Soiland-Reyes S, Dunlop I, Nenadic A, Fisher P, Bhagat J, Belhajjame K, Bacall F, Hardisty A, Nieva de la Hidalga A, Balcazar Vargas MP, Sufi S, Goble C (2013) The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Res 41:W557–W561
56. Oulas A, Karathanasis N, Louloupi A, Iliopoulos I, Kalantidis K, Poirazi P (2012) A new microRNA target prediction tool identifies a novel interaction of a putative miRNA with CCND2. RNA Biol 9:1196–1207

57. Coronnello C, Benos PV (2013) ComiR: combinatorial microRNA target prediction tool. Nucleic Acids Res 41:W159–W164
58. Bisognin A, Sales G, Coppe A, Bortoluzzi S, Romualdi C (2012) MAGIA(2): from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). Nucleic Acids Res 40:W13–W21
59. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human MicroRNA targets. PLoS Biol 2:e363
60. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microRNA target predictions. Nat Genet 37:495–500
61. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39:1278–1284
62. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J (2013) The RNAsnp web server: predicting SNP effects on local RNA secondary structure. Nucleic Acids Res 41:W475–W479
63. Halvorsen M, Martin JS, Broadaway S, Laederach A (2010) Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet 6:e1001074
64. Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder--a covariance model based RNA motif finding algorithm. Bioinformatics (Oxford, England) 22:445–452
65. Hiller M, Pudimat R, Busch A, Backofen R (2006) Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. Nucleic Acids Res 34:e117
66. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208
67. Rabani M, Kertesz M, Segal E (2008) Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. Proc Natl Acad Sci U S A 105:14885–14890
68. Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, Cristea IM, Tavazoie S (2012) Systematic discovery of structural elements governing stability of mammalian messenger RNAs. Nature 485: 264–268
69. Kazan H, Morris Q (2013) RBPmotif: a web server for the discovery of sequence and structure preferences of RNA-binding proteins. Nucleic Acids Res 41:W180–W186
70. Kazan H, Ray D, Chan ET, Hughes TR, Morris Q (2010) RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. PLoS Comput Biol 6:e1000832
71. Cereda M, Pozzoli U, Rot G, Juvan P, Schweitzer A, Clark T, Ule J (2014) RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. Genome Biol 15:R20
72. Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8:469–477

Chapter 2

A Computational Approach for the Discovery of Protein–RNA Networks

Domenica Marchese, Carmen Maria Livi, and Gian Gaetano Tartaglia

Abstract

Protein–RNA interactions play important roles in a wide variety of cellular processes, ranging from transcriptional and posttranscriptional regulation of genes to host defense against pathogens. In this chapter we present the computational approach *cat*RAPID to predict protein–RNA interactions and discuss how it could be used to find trends in ribonucleoprotein networks. We envisage that the combination of computational and experimental approaches will be crucial to unravel the role of coding and noncoding RNAs in protein networks.

Key words Protein–RNA interactions, Interaction prediction, Ribonucleoprotein networks, Messenger RNA, Noncoding RNA, *cat*RAPID

1 Introduction

The human genome harbors >1500 genes encoding proteins containing at least one RNA-binding domain (RBD) [1]. The number of proteins with identified RNA-binding ability (RBP), either possessing canonical or noncanonical RBDs [2, 3], is increasing. The fact that some proteins bind to transcripts through domains or regions that are not specifically evolved to this precise purpose [3, 4] is particularly intriguing. Indeed, recent manuscripts suggest a scenario where unexpected players can exert crucial functions in processes that were previously thought of as exclusively regulated by selected RBD-containing proteins [5].

Computational models represent an important source of information that can be exploited to identify hidden trends and understand the basics of molecular recognition. As a matter of fact, bioinformatics tools can perform exhaustive analyses and extract distinctive features, hence facilitating the design of new experiments. For example, it has been shown in several studies that the

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_2,

composition of primary protein structure, and the physicochemical properties associated with it, can be used to describe the amino acid regions that are more likely to be involved in binding to RNA molecules [6, 7]. Due to the limitations of current experimental approaches, it remains difficult to simultaneously investigate the plethora of RBPs bound to a single transcript and RNA regions that are likely to be involved in the binding. This has resulted in experimentalists having to rely on protein analysis to investigate specific signatures.

We developed an algorithm, *cat*RAPID, to investigate protein–RNA associations involved in regulatory mechanisms [8]. We trained *cat*RAPID on a large set of protein–RNA pairs available in the Protein Data Bank [9] to discriminate interacting and non-interacting molecules using the information contained in primary structures. *cat*RAPID relies on the ViennaRNA package [10], which has an accuracy of ~76 % [11], to generate predictions of secondary structure ensembles. These structures are then analyzed to extract information on the pairing profile of each nucleotide. By means of this procedure, the probability of *cat*RAPID predicting a protein–RNA interaction has a 72 % correlation with secondary structure information. However, a higher correlation factor is consistently expected with the enhancement of secondary structure prediction accuracies. As the predictive power of global RNA structure becomes less accurate as the length of the RNA increases [12], we developed the *cat*RAPID *fragments* module that exploits the RNALfold algorithm [11] to determine interactions for the most stable local structure.

2 *cat*RAPID Modules

The *cat*RAPID approach (http://s.tartaglialab.com/page/catrapid_group) [8, 13] has been developed to predict protein associations with coding and noncoding RNAs [14, 15] (Table 1). In our method, the contributions of secondary structure, hydrogen bonding, and van der Waals are combined together into the *interaction profile*:

$$\vec{\Phi}_x = \alpha_{\mathrm{H}} \vec{\mathrm{H}}_x + \alpha_{\mathrm{W}} \vec{\mathrm{W}}_x + \alpha_{\mathrm{S}} \vec{\mathrm{S}}_x$$

where the variable x indicates RNA ($x = \mathrm{r}$) or protein ($x = \mathrm{p}$). The hydrogen bonding profile, denoted by $\vec{\mathrm{H}}$, is the hydrogen bonding ability of each amino acid (or nucleotide) in a protein (or RNA) sequence:

$$\vec{\mathrm{H}} = \mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_{\mathrm{length}}$$

Table 1
Algorithms of the *cat*RAPID suite. Computational models, their applications and examples

Type of analysis	Algorithm	Result	Examples
The protein–RNA pair of interest are <750 aa and 1200 nt in length	*cat*RAPID *graphic* and *strength* modules	The score will provide the *propensity* to interact as well as an estimate of the *strength* of interaction	CSR system [13] FMRP [16]
Protein (or RNA) is larger than 750 aa (1200 nt)	*cat*RAPID *fragments* (*protein and RNA* option)	The *binding sites* of both molecules are visualized	SNCA [28] UNR (this work)
RNA is >10,000 nt and protein <750 aa	*cat*RAPID *fragments* (*long RNA* option)	The *binding sites* of the protein on the RNA sequence are identified	hnRNP-L [18] Xist [14]
Protein (transcript) partners of an RNA (protein) of interest	*cat*RAPID *omics*	*Propensity, strengths, binding motifs* are ranked in a table	HuR [19] LIN28B [19]
Interacting protein (transcript) partners co-expressed in human tissues	*cat*RAPID *omics express*	*Propensity, strengths, binding motifs* and *expression patterns* are characterized	TIA1 [18] MSI [18]

Similarly, $\vec{S}$ represents the secondary structure occupancy profile and $\overline{\overline{W}}$ the van der Waals' profile. The *interaction propensity* π is defined as the inner product between the protein propensity profile $\vec{\Psi}_p$ and the RNA propensity profile $\vec{\Psi}_r$ weighted by the *interaction matrix I*:

$$\pi = \vec{\Psi}_p I \vec{\Psi}_r$$

The algorithms to compute protein–RNA interactions are available at our group webpage http://service.tartaglialab.com/page/catrapid_group

2.1 catRAPID Graphic

Our original algorithm predicts the interaction propensity of a protein–RNA pair reporting the discriminative power DP, which is a measure of interaction strength with respect to the training sets [8]. The DP ranges from 0 % (the case of interest is predicted to be negative) to 100 % (the case of interest is a positive). DP values above 50 % indicate that the interaction is likely to take place, whereas DPs above 75 % represent high-confidence predictions. Due to computational requirements (intense CPU usage), the *cat*RAPID graphic algorithm accepts protein sequences with a length between 50 and 750 aa and RNA sequences between 50 and 1200 nt [15].

2.2 catRAPID Fragments

When input sequences exceed the length compatible with our computational requirements (i.e.: protein length > 750aa or RNA length > 1200 nt), the *cat*RAPID *graphic* cannot be used to calculate the interaction propensity [14, 15]. To overcome this limitation, we developed a procedure called *fragmentation*, which cuts polypeptide and nucleotide sequences into fragments followed by the prediction of the interaction propensities. Two types of fragmentation are possible:

- *Protein and RNA uniform fragmentation* (transcripts < 10,000 nt) [15]: The fragmentation approach is based on the division of protein and RNA sequences into 104 overlapping segments. This analysis is particularly useful to identify regions involved in the binding.
- *Long RNA* weighted fragmentation (for transcripts > 10,000 nt) [14]: The use of RNA fragments is introduced to identify RNA regions involved in protein binding. The RNALfold algorithm from Vienna package is employed to select RNA fragments in the range between 100 and 200 nt with predicted stable secondary structure.

2.3 catRAPID Strength

We previously observed that the strength correlates with chemical affinities [14], which suggests that the interaction propensity can be used to estimate the strength of association [16]. *cat*RAPID *strength* algorithm calculates the strength of a protein–RNA pair with respect to a reference set [13]. Random associations between polypeptide and nucleotide sequences are used to build the reference set. Since little interaction propensities are expected from random associations, the reference set is considered a negative control. Reference sequences have the same length as the pair of interest to guarantee that the interaction strength is independent of protein and RNA length. The interaction strength ranges from 0 % (non-interacting) to 100 % (interacting). Interaction strengths above 50 % indicate propensity to bind.

2.4 catRAPID Omics

The method is based on *cat*RAPID [8] algorithm and performs high-throughput predictions of protein–RNA interactions. *cat*RAPID *omics* enables: (1) the calculation of protein–RNA interactions on a large scale (up to 105 associations) in a reasonable time; (2) the submission of protein and RNA sequences without any length restriction; and (3) to focus on specific protein regions able to bind nucleic acid molecules [17] (Table 2).

- The time required by the original *cat*RAPID algorithm for predicting a single RNA–protein interaction strictly depends on the features of the input molecules, which are computed on the fly for each submission (using parallel calculation, <10 min are required for proteomic interactions of one RNA molecule).

Table 2
Composition of reference libraries used in *cat*RAPID *omics*

	Proteome					
	Full proteins		Domains		Transcriptome	
Model organisms	RNA	DNA	RNA	DNA	Coding	Noncoding
Caenorhabditis elegans	79	304	255	339	16613	8385
Danio rerio	82	323	311	391	21752	4589
Drosophila melanogaster	71	283	318	447	6307	1109
Homo sapiens	472	2152	1907	7432	105586	18553
Mus musculus	379	1518	1573	3073	42951	7243
Rattus norvegicus	168	592	689	902	13593	4823
Saccharomyces cerevisiae	261	389	508	431	3711	396
Xenopus tropicalis	70	184	279	253	2260	1278
Total	1582	5745	5840	13,268	98548	46376

Full-length (protein between 50 and 750 amino acids in length) and domains (derived from proteins >50 amino acids in length) are used as input of the method. Both sets are divided in additional groups, based on the ability of proteins to bind to RNA or DNA. Transcriptome searches use coding and noncoding RNAs, depending on the annotation in ENSEMBL version 68. The length of the transcripts in the datasets ranges from 50 to 1200 nucleotides, but longer RNAs can be added to the libraries

- To speed up the calculation of a far greater number of interactions, we introduce in *cat*RAPID *omics* a system of organism-specific feature libraries.

2.5 catRAPID Extensions

*cat*RAPID is interfaced with other methods to improve its predictive power [18]. Very recent implementations include the analysis of co-expression networks [19], the *clever*Suite approach to predict the RNA-binding ability of proteins [20] and the SeAMotE algorithm to identify regulatory elements coding/noncoding transcripts [21]:

- To train the *clever*Suite (http://s.tartaglialab.com/page/clever_suite), we focused on RNA-interacting proteins detected with UV cCL and PAR-CL protocols on proliferating HeLa cells followed by sequencing and compared them with the rest of cell lysate [3]. Analysis of physicochemical properties revealed a strong and consistent RNA binding property of the dataset (RNA-binding scales [3, 22, 23] discriminate 32–35 % of the entire database). The *clever*Suite selects the scales for nucleic acid binding [22, 23], membrane [24], burial [25] and aggregation [26] propensities, achieving a sensitivity of 0.72 and false positive rate of 0.24 on the entire dataset.

We applied the *clever*Suite to proteins that are classified as putative RNA-binding because they lack the canonical RNA-binding domains [3]. We observed correct classification associated with a sensitivity of 0.83 and false positive rate of 0.15, which indicates very high agreement with experimental data.

- Detection of regulatory motifs is a challenging task. For this reason, we developed the SeAMotE algorithm (http://s.tartaglialab.com/new_submission/seamote) [21], which provides an easy-to-use interface and allows the exhaustive analysis of large-scale datasets. Our approach offers unique features such as the discrimination based on the actual occurrences (i.e., pattern counts are not estimated) in the datasets, the choice of multiple reference backgrounds (shuffle, random, or custom) and the output of the most significant motifs in the whole span of tested motif widths, thus providing a wide range of solutions. In conclusion, our web-server is a powerful tool for the identification of enriched sequence patterns that characterize recognition process between proteins and nucleic acids. To evaluate SeAMotE performances on large-scale datasets, we collected recent CLIP experiments and assessed ability to identify significantly enriched motifs (Fisher's exact test). In each case analyzed, we compared RNAs bound to a specific protein (foreground set) with the same amount of non-interacting transcripts (background set). The DREME [27] algorithm was used as a reference to evaluate the performance of our system. Our method achieves both higher discrimination, which is the ability to separate the foreground from the background set, and significance, denoted by lower P-values associated with sequence motifs. In addition, SeAMoTe also shows very high sensitivity (~90 %) and accuracy (80 %).

3 *cat*RAPID Applications

3.1 Self-Regulatory Mechanisms Controlling Protein Production

We used the *cat*RAPID method to unravel self-regulatory pathways (autogenous interactions) controlling gene expression [15]. We discovered that aggregation-prone and structurally disordered proteins have a strong propensity to interact with their own mRNA [28]. Our results [15, 29] are in agreement with previous experimental work:

- It has been shown that the amyloidogenic TAR DNA binding protein 43 TDP-43 and *Fragile X* mental retardation protein FMRP interact with the 3′ UTR of their own mRNA to control protein production [15, 30, 31]. As overexpression leads to high protein concentration and enhanced amyloidogenicity [32, 33], it is possible that autogenous interactions prevent from generation of potentially toxic aggregates.

- The biosynthesis of tumor suppressor p53 is controlled by a translational autoregulatory feedback mechanism in which the p53 protein binds to its own mRNA in the 5′ terminal region, resulting in translational repression [34]. Indeed, it has been reported that naturally occurring mutations of p53 are associated with an increase of the aggregation potential [35]. In these regards, self-regulation of p53 can be seen as a way to control its aggregation potential.
- HSP70, the major stress-induced heat shock protein, regulates its own expression by interacting with its mRNA. Prolonged presence of HSP70 is detrimental for the cell, as it promotes aggregation. From ex vivo experiments, it has been shown that an increase in the degradation of HSP70 mRNA accompanies aggregation of HSP70 [36]. The interaction of HSP70 with its own mRNA (3′ UTR) suggests a self-limiting mechanism to reduce chaperone production and to avoid potential toxic effects in absence of stress [36].
- Moreover, the content of ribosomal proteins in eukaryotic cells is controlled by changes in the degradation rate of newly synthesized proteins. Such a high degree of coordination is achieved through the use of common regulatory elements in the genes and mRNAs of ribosomal proteins. In the majority of cases, regulation follows a feedback pattern, involving interactions of a ribosomal protein with its own pre-mRNA. This regulatory mechanism provides the required level of each individual ribosomal protein in the cell independently of other ribosomal proteins, which is crucial for extra-ribosomal functions. In the case of ribosomal proteins rpS26 and rpS13, high affinity for pre-mRNA fragments containing first introns has been found [37].

3.2 X-Chromosome Dosage Compensation

Dosage compensation of sex chromosomes equalizes expression of X-linked genes in organisms where males and females have a different number of X chromosomes. In mammals, Xist-mediated X chromosome inactivation (XCI) implies a complex network of macromolecular associations orchestrated by epigenetic modifiers as well as splicing and transcription factors.

- We used *cat*RAPID to investigate the interactions of the long noncoding Xist with Polycomb group proteins as well as YY1, SAF-A, ASF, and SATB1 proteins. In striking agreement with experimental evidence, we predicted protein binding sites and their affinities for Xist regions. We used our analysis to integrate the existing model of XCI into a new framework in which the transcriptional repressor YY1 tethers Xist to the X chromosome and nuclear matrix proteins SAF-A and SATB1 guide its translocation [14].

In *Drosophila melanogaster*, translational inhibition of male-specific *msl-2* messenger RNA by female-specific protein SXL is crucial for X-chromosome dosage compensation. Experimental studies identified an RNA-binding protein, UNR, as a fundamental co-repressor recruited by SXL to the 3′ UTR of *msl-2* mRNA for translation inhibition in females.

- RNA affinity chromatography and UV crosslinking assays show that UNR transcript and its 5′ UTR (nucleotides 1–261) efficiently bind to UNR protein, whereas 3′ UTR (nucleotides 261–447) does not [38]. Our calculations, carried out with *cat*RAPID fragments ("*Protein and RNA uniform fragmentation*" option), reproduce experimental results in great detail, identifying the cold shock domains (CSD; Fig. 1a) [39] involved

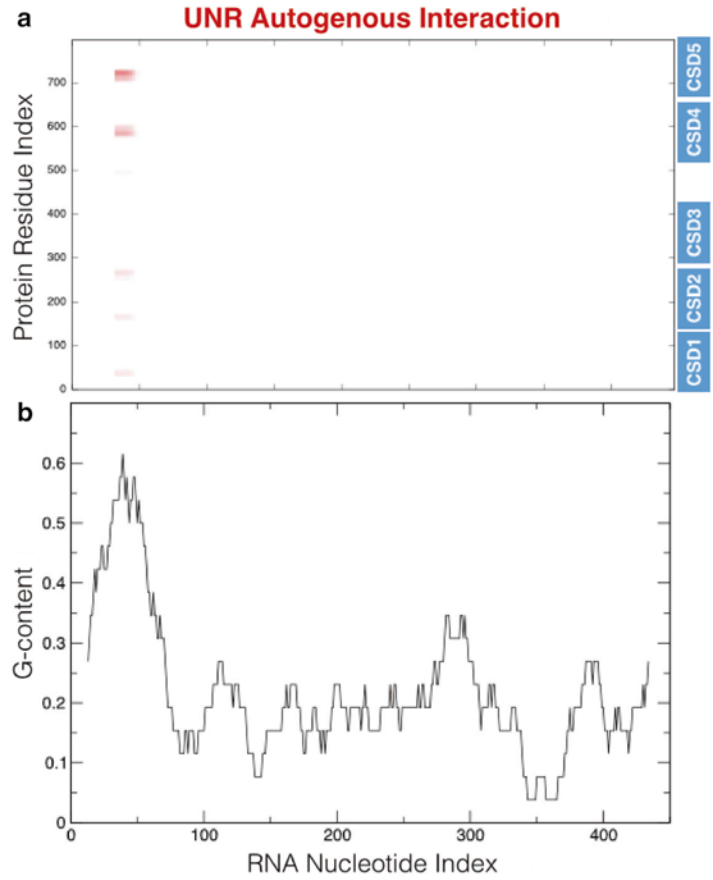

Fig. 1 UNR autogenous interactions. UNR transcript and its 5′ UTR (nucleotides 1–261) bind to UNR protein [38]. (**a**) Our calculations, carried out with *cat*RAPID fragments ("*Protein and RNA uniform fragmentation*" option), recapitulate experimental results in great detail, identifying cold shock domains (CSD) [39] involved in autogenous interaction; (**b**) In agreement with experimental evidence [38], UNR protein is predicted to bind to purine repeats, such as the guanine-rich region of UNR 5′ UTR

in autogenous interaction. In agreement with previous reports [38], UNR protein is predicted to bind to purine repeats, such as the guanine-rich region of its 5′ UTR (nucleotides 26–77; Fig. 1b).

3.3 Alternative Splicing

Recent studies indicate that nonsense-mediated decay (NMD) is an important element in alternative splicing regulation [40] and is associated with self-regulatory mechanisms:

- Polypyrimidine tract binding protein (PTB) regulates its own expression through a negative-feedback loop involving alternative splicing, which requires binding to mRNA and subsequent NMD triggered by exon skipping [41]. PTB autogenous interaction is particularly relevant because over-expression of the protein results in cell toxicity [42, 43].
- Similarly, heterogeneous nuclear ribonucleoprotein L hnRNP-L is able to induce NMD by associating with its mRNA [44]. Our predictions, carried out with *cat*RAPID fragments ("Long RNA" fragmentation option) indicate that hnRNP-L interacts with its own transcript in three different intronic regions located between exons 1–2, 6–7 and 9–10, which is in complete agreement with experimental evidence [44]. More specifically, we predict that hnRNP-L protein binds with high affinity to the 3′ CA cluster 6A of the hnRNP-L gene (intron 6) and not to sequence 6A (negative control), which is perfectly in agreement with the results of in vitro splicing assays performed by Rossbach et al. [44].

These and other results indicate that autogenous interactions occur in UTR/intronic regions and play a role in controlling protein production [28].

References

1. Ascano M, Gerstberger S, Tuschl T (2013) Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. Curr Opin Genet Dev 23:20–28
2. Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8:479–490
3. Castello A et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406
4. Baltz AG et al (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690
5. Kwon SC et al (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol 20:1122–1130
6. Terribilini M et al (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res 35: W578–W584
7. Fernandez M et al (2011) Prediction of dinucleotide-specific RNA-binding sites in proteins. BMC Bioinformatics 12(Suppl 13):S5
8. Bellucci M, Agostini F, Masin M, Tartaglia GG (2011) Predicting protein associations with long noncoding RNAs. Nat Methods 8: 444–445
9. Berman HM et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
10. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429–3431

11. Lorenz R et al (2011) ViennaRNA Package 2.0. Algorithms Mol Biol 6:26
12. Doshi KJ, Cannone JJ, Cobaugh CW, Gutell RR (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. BMC Bioinformatics 5:105
13. Cirillo D, Agostini F, Tartaglia GG (2013) Predictions of protein–RNA interactions. WIREs Comput Mol Sci 3:161–175
14. Agostini F, Cirillo D, Bolognesi B, Tartaglia GG (2013) X-inactivation: quantitative predictions of protein interactions in the Xist network. Nucleic Acids Res 41:e31
15. Cirillo D et al (2013) Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. RNA 19:129–140
16. Johnson R, Noble W, Tartaglia GG, Buckley NJ (2012) Neurodegeneration as an RNA disorder. Prog Neurobiol 99:293–315
17. Finn RD et al (2009) The Pfam protein families database. Nucleic Acids Res 38(Database): D211–D222
18. Cirillo D, Livi CM, Agostini F, Tartaglia GG (2014) Discovery of protein-RNA networks. Mol Biosyst 10:1632–1642
19. Cirillo D et al (2014) Constitutive patterns of gene expression regulated by RNA-binding proteins. Genome Biol 15:R13
20. Klus P et al (2014) The cleverSuite approach for protein characterization: predictions of structural properties, solubility, chaperone requirements and RNA-binding abilities. Bioinformatics 30:1601–1608
21. Agostini F, Cirillo D, Ponti RD, Tartaglia GG (2014) SeAMotE: a method for high-throughput motif discovery in nucleic acid sequences. BMC Genomics 15:925
22. Terribilini M et al (2006) Prediction of RNA binding sites in proteins from amino acid sequence. RNA 12:1450–1462
23. Lewis BA et al (2011) PRIDB: a protein–RNA interface database. Nucleic Acids Res 39:D277–D282
24. Nakashima H, Nishikawa K, Ooi T (1990) Distinct character in hydrophobicity of amino acid compositions of mitochondrial proteins. Proteins 8:173–178
25. Wertz DH, Scheraga HA (1978) Influence of water on protein structure. An analysis of the preferences of amino acid residues for the inside or outside and for specific conformations in a protein molecule. Macromolecules 11:9–15
26. Pawar AP et al (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350:379–392
27. Bailey TL (2011) DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27:1653–1659
28. Zanzoni A et al (2013) Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. Nucleic Acids Res 41(22): 9987–9998
29. Agostini F et al (2013) catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinformatics 29(22):2928–2930
30. Ayala YM et al (2011) TDP-43 regulates its mRNA levels through a negative feedback loop. EMBO J 30:277–288
31. Schaeffer C et al (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. EMBO J 20:4803–4813
32. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M (2007) Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem Sci 32:204–206
33. Baldwin AJ et al (2011) Metastability of native proteins and the phenomenon of amyloid formation. J Am Chem Soc 133:14160–14163
34. Ewen ME, Miller SJ (1996) p53 and translational control. Biochim Biophys Acta 1242: 181–184
35. Xu J et al (2011) Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. Nat Chem Biol 7:285–295
36. Balakrishnan K, De Maio A (2006) Heat shock protein 70 binds its own messenger ribonucleic acid as part of a gene expression self-limiting mechanism. Cell Stress Chaperones 11:44–50
37. Parakhnevitch NM, Ivanov AV, Malygin AA, Karpova GG (2007) Human ribosomal protein S13 inhibits splicing of its own pre-mRNA. Mol Biol 41:44–51
38. Schepens B et al (2007) A role for hnRNP C1/C2 and Unr in internal initiation of translation during mitosis. EMBO J 26:158–169
39. Triqueneaux G, Velten M, Franzon P, Dautry F, Jacquemin-Sablon H (1999) RNA binding specificity of Unr, a protein with five cold shock domains. Nucleic Acids Res 27:1926–1934
40. Kalyna M et al (2011) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res. doi:10.1093/nar/gkr932
41. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CWJ (2004) Autoregulation

of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell 13:91–100

42. Lin S, Wang MJ, Tseng K-Y (2013) Polypyrimidine tract-binding protein induces p19(Ink4d) expression and inhibits the proliferation of H1299 cells. PLoS One 8:e58227
43. Izquierdo JM et al (2005) Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Mol Cell 19:475–484
44. Rossbach O et al (2009) Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. Mol Cell Biol 29:1442–1451

Part II

Expression Studies

Chapter 3

Transcriptional Regulation with CRISPR/Cas9 Effectors in Mammalian Cells

Hannah Pham, Nicola A. Kearns, and René Maehr

Abstract

CRISPR/Cas9-based regulation of gene expression provides the scientific community with a new high-throughput tool to dissect the role of genes in molecular processes and cellular functions. Single-guide RNAs allow for recruitment of a nuclease-dead Cas9 protein and transcriptional Cas9-effector fusion proteins to specific genomic loci, thereby modulating gene expression. We describe the application of a CRISPR-Cas9 effector system from *Streptococcus pyogenes* for transcriptional regulation in mammalian cells resulting in activation or repression of transcription. We present methods for appropriate target site selection, sgRNA design, and delivery of dCas9 and dCas9-effector system components into cells through lentiviral transgenesis to modulate transcription.

Key words CRISPR/Cas9, Cas9-effectors, Transcriptional regulation, Gene activation, Gene repression

1 Introduction

Dissection of individual gene function through targeted downregulation or overexpression greatly benefits from versatile systems that allow for easy manipulation of target genes. Loss-of-function and gain-of-function high-throughput screens for factors involved in cellular function are also highly desirable. CRISPR (clustered regularly interspaced short palindromic repeats)-associated (Cas) systems offer the opportunity to manipulate endogenous genes at the level of transcriptional regulation, allowing for insight into the role of individual genes and gene regulatory networks in their endogenous context.

The CRISPR/Cas system present in bacteria and archaea serves as an adaptive immune system detecting and silencing foreign DNA [1, 2]. CRISPR systems incorporate invading DNA sequences into the genome at CRISPR loci that are later transcribed and used to guide nucleases to cleave invading DNA. Type II CRISPR systems require a single CRISPR-associated (Cas) gene, Cas9, a trans-activating CRISPR RNA (tracrRNA) and a

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_3,

CRISPR RNA (crRNA) to carry out this directed, sequence-specific DNA degradation [3, 4].

Type II CRISPR systems have recently been adapted for genome editing in mammalian cells, where a human codon-optimized version of the Cas9 protein introduced along with a crRNA and tracrRNA is capable of inducing site-specific double-strand breaks [5]. It is common to use a chimeric version of the crRNA and tracrRNA to form a single guide RNA (sgRNA), further simplifying the system to just two components consisting of the Cas9 protein and sgRNA to achieve site-specific recruitment and target cleavage [5, 6].

Target sites for CRISPR-mediated DNA cleavage are dependent on the presence of a specific protospacer adjacent motif (PAM) sequence 3′ to the targeted genomic DNA sequence [3, 7]. PAM sequences differ between CRISPR systems, for example, for the species *Streptococcus pyogenes* (Sp) the PAM sequence is "NGG" [7] whereas for *Neisseria meningitides* (Nm), the PAM sequence is "NNNNGATT" [8]. Since the Cas9 endonuclease cleaves target DNA only if the PAM sequence is present, the possible target sites are therefore defined by the CRISPR system used. However certain variations in the PAM sequences can be tolerated, including "NAG" for the Sp system and "NNNNGCTT" for the Nm system [9, 10], increasing the frequency of possible genomic target sites.

Several methods have been published describing the use of CRISPR-Cas9 systems for genome editing (e.g., [11]). However, since Cas9 can be guided to target specific genomic sequences, and effector proteins can be fused to Cas9, this programmable system is ideal for many applications beyond gene targeting through DNA cleavage. The protocol presented here focuses on CRISPR-based methods to manipulate gene regulation in mammalian cells using a Sp CRISPR-Cas9 system coupled to the Krueppel repressor associated box (KRAB) domain and the quadruple tandem repeat of the herpes simplex virus VP16 (VP64) transactivation domain [12–15]. In extension, the methods described can be easily transferred to CRISPR-Cas9 systems from other species as well as to Sp Cas9 coupled to different effectors.

In order to utilize the CRISPR system to affect gene regulation, it is first necessary to inactivate the DNA cleavage activity of the Cas9 protein. Cas9 contains two nuclease domains, a RuvC-like domain and a HNH domain, the activities of which are responsible for performing the double strand break once the Cas complex is recruited to DNA [3]. Inactivation of Cas9 nuclease activity is therefore commonly achieved through mutation of key catalytic residues in both of the Cas9 nuclease domains (Sp system: amino acid changes D10A and H840A) creating a nuclease dead version of the Cas9 protein (dCas9) [3, 16]. dCas9 can then be used with an sgRNA to enable recruitment to specific genomic loci without inducing DNA cleavage.

sgRNA-directed dCas9 recruitment to DNA has been shown to interfere with gene transcription (CRISPRi) through steric hindrance by preventing transcriptional elongation, polymerase binding or transcription factor binding thereby influencing gene expression levels [16]. Recently, additional tools have been developed for the manipulation of gene regulation through fusion of dCas9 to effector domains as have previously been developed for TALE and zinc finger systems [17]. Coupling the dCas9 to effectors allows for site-specific delivery of any effector domain of interest, and many different effector domains have been successfully fused to dCas9 systems to modulate gene expression in mammalian cells. Positioned by the sgRNA-dCas9 complex, the effector domain can dictate the cellular response resulting in gene repression (CRISPRi) (e.g., KRAB, SID domains) or gene activation (CRISPRa) (e.g., VP16, VP64, p65AD domains) when targeted to regions upstream of transcriptional start sites [16, 18–25].

dCas9 effector-mediated gene regulation can be applied to a broad range of studies including influencing cellular states in stem cells through differentiation [24] to dissecting of the contribution of individual genomic elements or transcription factor binding sites to the regulation of a single gene. Critically, due to the speed and ease of sgRNA design, high-throughput approaches have been developed for gain- or loss-of-function screens in mammalian cells [26, 27].

Here, we describe a stepwise protocol for applying a CRISPR-Cas9 system from *Streptococcus pyogenes* for gene regulation in mammalian cells through a KRAB repressor domain or VP64 activation domain. The protocol includes preparation of a dCas9-effector system, design and cloning of sgRNAs (Subheadings 3.1–3.3) in addition to lentiviral production and transgenesis (Subheadings 3.4 and 3.5). These steps can be applied to other autologous CRISPR-Cas9 systems and can be utilized in a variety of cell types.

2 Materials

2.1 Plasmids

1. dCas9 and dCas9-effector plasmids
   pHAGE TRE-dCas9 (Addgene #50915)
   pHAGE TRE-dCas9-VP64 (Addgene #50916)
   pHAGE TRE-dCas9-KRAB (Addgene #50917)
   pHAGE EF1alpha-dCas9-VP64 (Addgene #50918)
   pHAGE EF1alpha-dCas9-KRAB (Addgene #50919)
2. sgRNA plasmids
   pLenti Sp BsmBI sgRNA Puro (Addgene #62207)
   pLenti Sp BsmBI sgRNA Hygro (Addgene #62205)
3. Lentiviral plasmids

pHDM-G (DNASU #235)
pHDM-Hgpm2 (DNASU #236)
pHDM-tat1b (DNASU #237)
pRC/CMV-rev1b (DNASU #246)

2.2 Molecular Biology Reagents

1. DNA oligomers (25 nmol standard desalting conditions).
2. Molecular biology grade water.
3. ATP.
4. T4 polynucleotide kinase 10 U/μl.
5. BsmBI 10 U/μl.
6. Tris–acetate–EDTA (TAE) buffer.
7. Agarose.
8. QIAquick Gel Extraction Kit (Qiagen).
9. T4 DNA ligase 400 U/μl.
10. Stbl3 chemically competent *E. coli*.
11. Ampicillin sodium salt.
12. Luria broth (LB).
13. Luria agar.
14. QIAprep Spin Miniprep Kit (Qiagen).
15. EcoRI 20 U/μl.
16. QIAGEN Plasmid Maxi Kit (Qiagen).

2.3 Cell Culture Reagents

1. Cell line: human embryonic kidney (HEK) 293T/17 (ATCC # CRL-11268).
2. Dulbecco's Modified Eagle's Medium (DMEM) high glucose (4.5 g/L).
3. Fetal bovine serum (FBS).
4. GlutaMAX.
5. Nonessential amino acids (NEAA).
6. Sodium pyruvate.
7. 0.25 % trypsin.
8. 10 cm dish.
9. 6-well plate.
10. TransIT-293 transfection reagent (Mirus).
11. OptiMEM medium.
12. 10 ml syringes.
13. Millex-HV Syringe Filter Unit, 0.45 μm, PVDF, 33 mm.
14. Lenti-X concentrator.
15. Puromycin 10 mg/ml.
16. G418 50 mg/ml.

17. Hygromycin B 50 mg/ml.
18. Doxycycline.
19. Phosphate buffered saline (PBS).
20. Formalin.
21. Trypan Blue.
22. PBST (1× PBS + 0.2 % Triton 100).
23. Donkey serum.
24. Rat anti-hemagglutinin (HA) antibody (3F10) (Roche).
25. Secondary antibody.
26. Hoechst 33342 10 mg/ml.
27. Target specific qPCR primers and/or antibody.

3 Methods

3.1 Preparation of dCas9 and dCas9-Effector and sgRNA Plasmids

1. Obtain bacterial stocks of Sp dCas9 or dCas9-effector (KRAB and VP64) lentiviral plasmids available through Addgene. These plasmids are ready for use and are available as either constitutive or inducible versions (*see* **Note 1**).
2. Streak bacteria onto 100 μg/mL ampicillin LB-agar plates and grow overnight at 37 °C.
3. Pick a single colony and grow for 6–8 h at 37 °C with shaking at 250 rpm in 5 mL of LB media (containing 100 μg/mL ampicillin).
4. Transfer the 5 mL of culture to 250 mL of fresh LB media (containing 100 μg/mL ampicillin) and grow overnight with shaking at 250 rpm at 37 °C.
5. Prepare the plasmid DNA using the Qiagen Maxiprep kit according to the manufacturer's protocol for subsequent preparation in Subheading 3.3.
6. (Optional) Confirm plasmid integrity by diagnostic digest.
   (a) Digest 1 μg of plasmid DNA with 1 μL of each of the restriction enzymes indicated in Table 1 for 1 h at 37 °C.
   (b) Run the digest on a 1 % agarose gel to resolve the indicated band sizes

3.2 sgRNA Design

Guidelines for selecting genomic targets are discussed in **Note 2**. Since the efficiency of different sgRNAs in downstream applications may vary, we recommend designing multiple sgRNAs per target.

3.2.1 Designing DNA Oligomers for Sp sgRNA Cloning Manually

1. Search for the PAM sequence (NGG) within the region of interest to identify possible target sites with the following genomic context $GN_{19}NGG$. Since the sgRNA is expressed from a human U6 promoter, a G is required at the start of the sequence for expression.

Table 1
List of plasmids and respective restriction enzymes

Plasmid	Enzymes to digest with	Expected band sizes
pHAGE TRE-dCas9 (Addgene #50915)	AflIII	5919, 4385, 3953, 176
pHAGE TRE-dCas9-VP64 (Addgene #50916)	AflIII	5919, 4553, 3953, 176
pHAGE TRE-dCas9-KRAB (Addgene #50917)	AflIII	5919, 3953, 2760, 1824, 176, 17
pHAGE EF1alpha-dCas9-VP64 (Addgene #50918)	AflIII	6142, 3583, 3081
pHAGE EF1alpha-dCas9-KRAB (Addgene #50919)	AflIII	4355, 3583, 3081, 1824, 17
pLenti Sp BsmBI sgRNA Puro (Addgene #62207)	EcoRI, XhoI	7044, 1440
pLenti Sp BsmBI sgRNA Hygro (Addgene #62205)	XhoI	7743, 1550

2. BLAST the genomic context ($GN_{19}NGG$) to confirm that there are no identical sequences in your genome.
3. Target site sequences with palindromes or poly-T -G or -A runs [26] should be avoided.
4. Order the following oligomers for sgRNA cloning into either sgRNA backbone vector. Be sure to omit the PAM sequence (NGG) and to include the overhang sequences to facilitate cloning into the BsmBI sites in the Addgene #62205 or #62207 sgRNA backbones (*see* Fig. 1).
   (a) Forward oligo: ACACC (GN_{19}) G
   (b) Reverse oligo: AAAAC (N_{19} complement C) G

3.2.2 Designing DNA Oligomers for Sp sgRNA Cloning Online

1. Many sgRNA design tools are now publically available to facilitate sgRNA design [9, 28–31]. DNA regions of interest can be fed into these tools to identify genomic targets predicted to have minimal off-target effects and to generate the target site sequences for a variety of mammalian species. It is important to ensure that the sgRNAs are designed for the Sp Cas9 system as sgRNAs designed for orthogonal CRISPR systems will have different PAM sequences and therefore would not be compatible.
2. In order to clone target site sequences into the Addgene #62205 or #62207 sgRNA backbones it is necessary to order oligomers containing the target site sequence (GN_{19}) and to include the following overhang sequences:
   (a) Forward oligo: ACACC (GN_{19}) G
   (b) Reverse oligo: AAAAC (N_{19} complement C) G

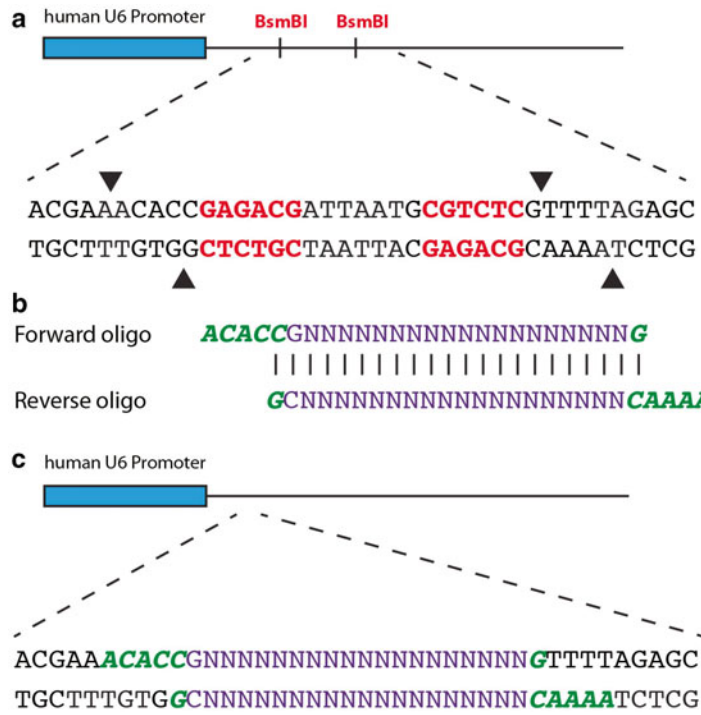

Fig. 1 Schematics for inserting annealed DNA oligomers into sgRNA expression backbones (Addgene #62207 and #62205). (**a**) Nucleotide sequence of backbone cloning region with BsmBI sites indicated in *bold red*. Digestion of the backbone with BsmBI results in DNA cuts at the positions indicated by *arrows*. (**b**) Annealed DNA oligomers containing the sgRNA target sequence. The overhangs for ligation are indicated in *bold green italics*. (**c**) Nucleotide sequence of ligation product containing oligomers inserted into sgRNA expression backbone

3.3 Cloning sgRNA into Lentiviral Expression Vector

1. Resuspend forward and reverse DNA oligomers for sgRNA cloning with molecular grade water at 100 μM. Prepare the following reaction to phosphorylate the oligomers and anneal them into dsDNA fragments for ligation into an sgRNA backbone.

100 μM Forward Oligo	2 μL
100 μM Reverse Oligo	2 μL
10× T4 polynuclease kinase buffer	2 μL
10 mM ATP	1 μL
T4 poly nuclease kinase	1 μL
Molecular grade water	12 μL
Total volume	20 μl

2. Incubate the reaction at 37 °C for 30 min.
3. Add 1 μL of 5 M NaCl and 29 μL Tris–EDTA (TE) Buffer

Heat the reaction in a heat block for 5 min at 95 °C, remove the block from the heat source and allow to cool slowly to room temperature. Alternatively, put in thermocycler, heat for 5 min at 95 °C and ramp to 25 °C at 0.1 °C/s

4. Dilute 1 μl of annealed sgRNA oligomers into 99 μl of TE buffer to prepare sample for ligation
5. Digest 2 μg of sgRNA backbone with 1 μL BsmBI overnight at 55 °C (*see* **Note 3**).
6. Run on a 1 % TAE agarose gel to resolve the 8500 bp fragment and detect potential uncut coiled plasmid
7. Gel purify the 8500 bp fragment using Qiagen Gel Extraction Kit
8. Elute in 30 μL elution buffer (EB) supplied with the Qiagen Gel Extraction Kit
9. Set up the following ligation overnight at 16 °C. Set up a ligation with no insert (annealed sgRNA oligomers) as a negative control.

Digested sgRNA backbone (from **step 8**)	100 ng
Annealed diluted sgRNA oligomers (from **step 4**)	4 μL
10× T4 DNA ligase buffer	2 μL
Molecular grade water	to 19 μL
T4 DNA ligase	1 μL

10. Transform 2 μL of ligation into Stbl3 competent cells according to manufacturer's protocol. The use of Stbl3 cells is critical, *see* **Note 4**.
11. Plate the transformed cells onto 100 μg/mL ampicillin LB-agar plates and grow overnight at 37 °C.
12. Check for colonies the following day. The negative control plate should have at least 10× fewer colonies than the plates where you have ligated in the annealed DNA oligomers. If the colony numbers are similar, this could indicate incomplete digestion of the sgRNA backbone.
13. Pick 4–6 colonies per ligation and grow overnight with shaking at 37 °C in 5 mL of LB culture (containing 100 μg/mL ampicillin).
14. Prepare the plasmid DNA using the Qiagen Miniprep kit according to the manufacturer's protocol
15. Perform a diagnostic restriction digestion to screen for the insertion of the annealed oligomers. Use the parental vector from **step 5**, Subheading 3.1 as a negative control.

(a) Digest 1 μg plasmid DNA with 1 μl EcoRI-HF in NEB Buffer 2 at 37 °C for 1–2 h.

(b) Add 1 μl BsmBI to the reaction and incubate at 55 °C for 1–2 h.

(c) Run the digest on a 1 % agarose gel until 5500 and 3000 bp fragments are resolved in the negative control. If an insert is present, the BsmBI sites will have been removed and there will be a single 8500 bp fragment due to linearization with EcoRI.

16. (Optional) To verify cloning and to confirm the target site sequence, Sanger sequence the insert with a U6 forward sequencing primer (GACTATCATATGCTTACCGT)
17. (Optional) Maxiprep verified clones for subsequent viral production using the Qiagen Maxiprep kit according to the manufacturer's protocol

3.4 Lentivirus Production

Both dCas9/dCas9-effector and sgRNA plasmids are third generation lentiviral vectors and can be packaged into lentiviral particles enabling delivery of both components of the dCas9 system into any cell type of interest.

3.4.1 Lentivirus Production

1. HEK293T/17 cells are maintained in DMEM supplemented with 10 % FBS, 1 % GlutaMAX, 1 % NEAA, and 1 % sodium pyruvate. Cells are passaged every 3–4 days at around 70 % confluence. Do not allow the cells to become confluent.
2. The day before transfection, passage the cells with 0.25 % trypsin, and plate at 1.3×10^5 cells/cm^2 or six million cells for each 10 cm dish. For alternatives *see* **Note 5**.
3. 2 h before transfection, feed the cells with fresh media
4. For the transfection, pipette DNA (dCas9-effector or sgRNA plasmids with packaging plasmids) in the following ratio 20:2:1:1:1 (Expression plasmid: pHDM-G: pHDM-Hgpm2: pHDM-tat1b: pRC/CMV-rev1b)

Plasmid	Desired (ng)
dCas9 or sgRNA	12,000
pHDM-G (DNASU #235)	1200
pHDM-Hgpm2 (DNASU #236)	600
pHDM-tat1b (DNASU #237)	600
pRC/CMV-rev1b (DNASU #246)	600
Total DNA	15,000

5. Transfect the cells with TransIT-293 transfection reagent in Opti-MEM according to manufacturer's protocol.

(a) Dilute the DNA in 1 mL of OptiMEM in a 1.5 mL centrifuge tube

(b) Add 45 μL of TransIT-293 transfection reagent to the center of the tube

(c) Mix gently using a 1 mL pipette

(d) Let incubate at room temperature, undisturbed, for 20 min

(e) Plate drop-wise onto cells and gently shake the plate to distribute the transfection mixture

6. The following day change the media on the transfected cells
7. 48 h after transfection, harvest the virus by passing the media through a 0.45 μM filter. Store the collected virus at −80 °C.
8. (Optional) Feed the cells with an additional 10 mL of fresh media and harvest the virus 24 h later repeating **step 7**.
9. (Optional) To obtain higher viral titers, virus can be concentrated immediately after harvest using Lenti-X concentrator according to the manufacturer's protocol. Alternatively, virus can be concentrated by centrifuging at 100,000 × *g* for 90 min at 4 °C.

3.4.2 Viral Titer: Calculating Multiplicity of Infection (MOI)

1. Maintain HEK293T/17 cells as in Subheading 3.4.1
2. The day before infection plate 5×10^4 HEK293T/17 cells/well onto a 6-well plate.
3. The day of infection, thaw lentivirus stock on ice, and change the media on the cells. Add 10 or 1 μL of virus to each well; be sure to leave one well uninfected as a negative control
4. 24 h after infection, change the media on infected cells
5. 48 h after infection, passage the cells at various densities (e.g., 1:10 and 1:100) onto a 10-cm dish and add selection reagent (depending on the lentiviral vector used this may vary, *see* **Note 6**)
6. Allow cells to grow until there are no cells left on the negative control plate and cells are large enough to visualize (usually 5–7 days), changing the media supplemented with selection reagent every 2–3 days
7. Carefully wash cells with PBS and fix with formalin solution for 30 min at room temperature. Wash with PBS, then stain with 2 mL of trypan blue for 10 min, and wash 2× with PBS. Count the number of colonies and calculate the viral titer:

$$\text{Titer}\left(\text{viral units / ml}\right) = \frac{\left(\#\text{colonies after selection}\right)}{\left(\text{Dilution factor from } \textbf{step 5} \text{ e.g. } 1/100\right)\left(\text{volume virus added from } \textbf{step 3} / 1000\right)}$$

8. Expected titers for sgRNAs are 10^6 to 10^7 viral particles/ml, and for dCas9-effectors are ~10^4 viral particles/ml

3.5 Virus Delivery/ Functional Assay

To determine the downstream effects of sgRNA-guided Cas9-effectors we commonly monitor cellular RNA and protein levels of target gene. Experiments should be designed to include negative controls, these could include sgRNAs designed to target a gene desert region or a sequence not present in the genome of interest. In the case of effector-based repression, it may be beneficial to include a dCas9 without a fused effector to distinguish between effector-mediated and steric hindrance effects.

3.5.1 Generation of Stable dCas9, dCas9-Effector Cell Lines

1. Passage desired cell type.
2. Incubate the cells with the concentrated dCas9 or dCas9-effector viruses.
3. 24 h after infection, change the media following lentiviral handling safety procedures.
4. 48 h after infection, change the media following lentiviral handling safety procedures and start selection to generate stable cell lines (*see* **Notes 6–8**). The required amount of time for selection will vary with cell line and selection cassette.
5. Following completion of selection, lines may be subcloned or a pool of stably transduced cells can used for future experiments.
6. (Optional) Expression of dCas9 or dCas9-effector can be confirmed through immunofluoresence detection of the HA-epitope tag. If doxycycline-inducible dCas9-effectors are used, cells must be maintained in 2 μg/ml doxycycline.
   (a) Wash cells with PBS, then fix in formalin solution for 30 min at room temperature.
   (b) Wash 1× with PBS, then 1× with PBST.
   (c) Incubate with blocking buffer (5 % donkey serum in PBST) for 30 min at room temperature
   (d) Dilute HA antibody in blocking buffer and stain for 3 h at room temperature. The HA antibody dilution should be optimized by lot.
   (e) Wash 3× PBST and incubate with appropriate secondary antibody for 2 h at room temperature.
   (f) Wash 3× PBST and incubate with 4 μg/ml Hoechst for 5 min at room temperature.
   (g) Wash 1× with PBS and visualize on microscope.

3.5.2 Transduction of sgRNA Lentivirus

1. Passage stably transduced dCas9 and dCas9-effector cell lines.
2. Incubate the cells with sgRNA lentiviruses. We usually infect cells with an MOI of ~1.
3. 24 h after infection, change the media following lentiviral handling safety procedures.

4. 48 h after infection, change the media following lentiviral handling safety procedures and start puromycin (for plasmid #62207) or hygromycin (for plasmid #62205) selection if desired.

3.5.3 Assessment of sgRNA Function

1. Following sgRNA transduction, successful CRISPRi or CRISPRa will result in relative loss or gain of target gene transcripts. Effects on gene expression can be assessed by quantitative PCR methods comparing test sgRNAs to negative control sgRNAs. The time required for the dCas9 or dCas9-effector to take effect may vary and a timecourse of the functional readout should be established for each new cell type.

4 Notes

1. dCas9, dCas9-KRAB and dCas9-VP64 plasmids are available through Addgene as doxycycline-inducible versions with a neomycin selection cassette (#50915 and #50916) and constitutively active versions with a puromycin selection cassette (#50917, #50918, #50919). Additional dCas9-effector systems can also be acquired from Addgene or principle investigators. If using the inducible version dCas9, doxycycline must be added to the media at 2 μg/ml after infection with dCas9 lentivirus to induce dCas9 expression. All vectors contain the dCas9 or dCas9-effector followed by a HA-epitope tag which can be used to detect expression of the dCas9. sgRNA plasmids for cloning are available with either a puromycin (#62207) or a hygromycin (#62205) selection cassette. Any of the dCas9 plasmids can be used with either of the sgRNA plasmids, however using a dCas9 lentiviral vector containing a selection cassette different from the sgRNA lentiviral vector will allow for selection of both components in a single cell.
2. Choosing an appropriate genomic target depends on the effector being used. When using dCas9 alone (without effector) to interfere with gene expression, genomic targets should either overlap or be downstream of the transcriptional start site. Successful CRISPRi effector mediated repression has been achieved by targeting dCas9KRAB within a −200 to +300 bp window of the transcriptional site [26, 32]. Targeting dCas9KRAB downstream of transcriptional start sites may result in a stronger repressive effect due to the combination of both KRAB-mediated repression and dCas9 mediated steric hindrance [26]. For CRISPRa, recent studies indicate that target sites within a −400 to +1 bp window of the transcriptional start site is optimal for dCas9-VP64 gene mediated activation [26,

32]. Designing multiple sgRNAs against a target and thereby recruiting multiple dCas9 or dCas9-effectors to the same region can increase the overall levels of activation [19–21, 24].

3. Esp3I (10U/μl), is an isoschizomer of BsmBI and can also be used to prepare the sgRNA backbone in place of BsmBI in Subheading 3.3, **step 5**. Digest (0.5 μL/μg of plasmid) overnight at 37 °C.
4. Due to the highly repetitive sequences found in lentiviral vectors (the repeat sequences in the LTR), Stbl3 cells are preferred for cloning. These cells reduce the rate of recombination caused by these repetitive sequences and conserve the integrity of the plasmid over time. The DNA yield is also found to be higher in these cells.
5. The CRISPR-effector system can be used for screening purposes and the lentiviral production can be scaled down into a 96-well format using 100 ng of total DNA per well for transfection (scaled from 15 μg of total DNA in **step 2**, subheading 3.4.1). In our experience, precoating plates with 5 μg/ml fibronectin prior to plating the HEK293T cells for viral production will allow for comparable virus titers between large and small well sizes. For high throughput virus production we do not filter the virus but rather freeze directly after harvesting.
6. The required dose and timing of selection reagents will vary with cell line. Typical timing and dose of common selection agents for mammalian cells are indicated but dose and concentration need to be optimized by cell type:

Concentration of selection reagent	Time required to complete selection
0.5–2 μg/ml Puromycin	2 days
25–50 μg/ml Hygromycin	4 days
50–300 μg/ml G418 (Neomycin)	6 days

If each component cannot be selected for (e.g., when transducing multiple sgRNAs or using a puromycin resistant dCas9 plasmid with a puromycin resistant sgRNA plasmid) then the MOI in Subheading 3.5.2, **step 2** can be increased to attain cells expressing all components.

7. Some cell lines are difficult to transduce and may benefit from a pre-incubation step with lentiviral particles before plating (Subheading 3.5). For mouse and human ESCs, we perform a 3-h incubation with virus in low attachment plates containing a minimal amount of media, prior to plating for continued culture. Alternatively, for some cell lines transduction efficiency

may be improved by the addition of polybrene during transduction.

8. Due to the large size of some dCas9-effector constructs, it may be technically challenging to generate high titers of dCas9-effector lentivirus. If avoidance of lentiviral delivery of dCas9-effectors is desirable, cotransfection of dCas9-effectors together with sgRNAs is sufficient to modulate gene expression [19]. Alternatively, stable dCas9-effector lines can be generated through random or target integration of a linearized plasmid. Continuous selection can be used to ensure maintained expression of the dCas9-effectors. The sgRNAs can then be delivered to these cell lines via lentiviral infection or by transfection.

Acknowledgements

This work was supported by The Leona M. and Harry B. Helmsley Charitable Trust (2015PG-T1D057) and a Charles H. Hood Foundation Child Health Research Award.

References

1. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167–170
2. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45:273–297
3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821
4. Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67–71
5. 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 339:819–823
6. 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 339:823–826
7. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading, Engl) 155:733–740
8. Zhang Y, Heidrich N, Ampattu BJ, Gunderson CW, Seifert HS, Schoen C, Vogel J, Sontheimer EJ (2013) Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. Mol Cell 50:488–503
9. 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. doi:10.1038/nbt.2647
10. 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 10:1116–1121
11. Yang L, Mali P, Kim-Kiselak C, Church G (2014) CRISPR-Cas-mediated targeted genome editing in human cells. Methods Mol Biol 1114:245–267
12. Urrutia R (2003) KRAB-containing zinc-finger repressor proteins. Genome Biol 4:231
13. Hirai H, Tani T, Kikyo N (2010) Structure and functions of powerful transactivators: VP16, MyoD and FoxA. Int J Dev Biol 54:1589–1596
14. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991) The evolution-

arily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc Natl Acad Sci U S A 88: 3608–3612

15. Campbell MEM, Palfreyman JW, Preston CM (1984) Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. J Mol Biol 180: 1–19
16. 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 152:1173–1183
17. Gersbach CA, Perez-Pinera P (2014) Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. Expert Opin Ther Targets 18:835–839
18. 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 154:442–451
19. 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 10:973–976
20. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK (2013) CRISPR RNA-guided activation of endogenous human genes. Nat Methods. doi:10.1038/nmeth.2598
21. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalila CS, Dadon DB, Jaenisch R (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. doi:10.1038/cr.2013.122
22. Farzadfard F, Perli SD, Lu TK (2013) Tunable and multi-functional eukaryotic transcription factors based on CRISPR/Cas. ACS Synth Biol. doi:10.1021/sb400081r
23. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F (2013) Optical control of mammalian endogenous transcription and epigenetic states. Nature 500: 472–476
24. Kearns NA, Genga RMJ, Enuameh MS, Garber M, Wolfe SA, Maehr R (2014) Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development 141:219–223
25. Hu J, Lei Y, Wong W-K, Liu S, Lee K-C, He X, You W, Zhou R, Guo J-T, Chen X, Peng X, Sun H, Huang H, Zhao H, Feng B (2014) Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors. Nucleic Acids Res 42:4375–4390
26. 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. doi:10.1016/j.cell.2014.09.029
27. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159:635–646
28. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 32:1262–1267
29. Zhu LJ, Holmes BR, Aronin N, Brodsky MH (2014) CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PLoS One 9:e108424
30. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L, Wang L, Hodgkins A, Iyer V, Huang X, Skarnes WC (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat Methods 11:399–402
31. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E (2014) CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res 42:W401–W407
32. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2014) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. doi:10.1038/nature14136

Chapter 4

Studying the Translatome with Polysome Profiling

Paola Zuccotti and Angelika Modelska

Abstract

Polysome fractionation by sucrose density gradient centrifugation followed by analysis of RNA and protein is a technique that allows to understand the changes in translation of individual mRNAs as well as genome-wide effects on the translatome. Here, we describe the polysome profiling technique and RNA as well as protein isolation procedures from sucrose fractions.

Key words Polysome, Profiling, Transcriptome, Translational control, Translatome, Sucrose gradient, RNA isolation, Protein isolation

1 Introduction

Polysome fractionation by sucrose density gradient centrifugation followed by analysis of RNA and protein allows to understand the translatome of normal, stress, and diseased states of various cell models as well as to study the changes of proteins that are present in or associated with polysomes. Since the development of this technique more than 40 years ago [1], polysome profiling has been extensively used to investigate protein translation and its dysregulation [2–4]. Transcripts have very complex lives and their fate is determined by a huge number of factors. Some of them might not be immediately destined for translation and might be for instance stored in P-bodies [5]. Therefore, measuring global mRNA levels is a rather poor approximation of the translational state of the cell [6], and more complex techniques such as polysome profiling are necessary to address fundamental questions about protein translation.

After stalling ribosomes on transcripts with cycloheximide, the cells are lysed and the lysate is applied onto a sucrose gradient. Free, monosome- and polysome-associated transcripts are then separated by ultracentrifugation and fractions are collected, from which RNA or protein can then be extracted for further analysis (Fig. 1a). Traditionally, the distribution pattern of specific mRNAs has been determined by northern blotting, but with recent

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_4,

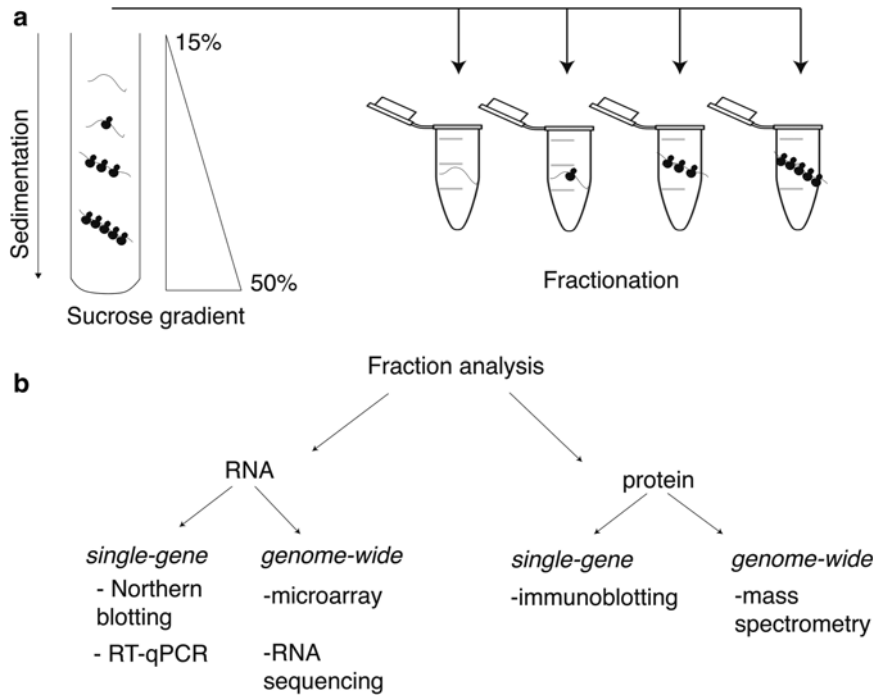

Fig. 1 (**a**) Separation of cell lysate by velocity sedimentation in a sucrose gradient on the basis of their differing "effective" sizes, or, roughly, on the basis of molecular weights. After centrifugation, fractions are collected under continuous reading of optical density at 254 nm. Free RNAs and other low-molecular-weight components of the cytoplasm are found at the top and polysomes are found at the bottom of the gradient. (**b**) Downstream processing of RNA and proteins obtained from polysome profiling. The most common techniques used are indicated. RT-qPCR—reverse transcription-quantitative PCR

technological advancements nowadays mostly reverse transcription (RT)-qPCR, microarray, and RNA sequencing technologies are used (Figs. 1b and 2a). Equally, proteins interacting with polysomes or ribosomal proteins themselves can be studied using immunoblotting or high-throughput technologies such as mass spectrometry (Figs. 1b and 2b).

From the mRNA distribution pattern obtained, it is then possible to determine the efficiency with which the transcripts are recruited to the translational machinery (by assaying the percentage of transcripts being associated with ribosomes, i.e., the ribosomal occupancy). Another useful parameter is the ribosomal density, which assesses the number of ribosomes with which the mRNA is associated [7]. Both of these parameters allow to derive the *translational efficiency* for different transcripts. Therefore, it is possible to study directly the effects on translation of individual transcripts upon the perturbation of the system, for example growth factor treatment, serum starvation, various differentiation stages, and over- or under-expression of selected genes [4, 8–11].

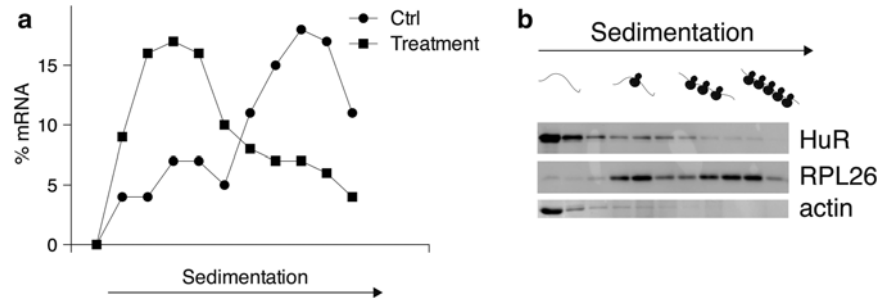

Fig. 2 (**a**) Relative distribution of a transcript along single fractions of the polysome profile with/without a treatment influencing its translation. (**b**) Immunoblotting showing the distribution of a protein associated with polysomes (HuR), a ribosomal protein (RPL26), and a protein not associated with polysomes (actin)

2 Materials

Prepare all solutions using diethyl pyrocarbonate (DEPC) ultrapure water and analytical grade reagents. Diligently follow all waste disposal regulations when disposing waste materials.

DEPC water, 1 % DEPC: Add 1 mL DEPC to 1000 mL ultrapure water, stirring overnight, autoclave (*see* **Note 1**).

2.1 Preparation of Cell Lysate

1. Tissue culture dishes (100 mm), regular media, and cell culture equipment.
2. 10 mg/mL cycloheximide in DEPC water: Add 100 mg in 10 mL DEPC water, vortex well to dissolve. Aliquot and store at −20 °C.
3. Lysis buffer: 10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris–HCl pH 7.5, 1 % Triton X-100, 1 % sodium deoxycholate, 0.2 U/μL RNase inhibitor, 1 mM DTT, 0.1 mg/mL cycloheximide (*see* **Note 2**).

   Stock Solutions for Lysis Buffer:

   10× Sodium deoxycholate: Add 1 g of sodium deoxycholate in 10 mL DEPC water (10 % w/v final). Store at room temperature.

   10× Triton X-100: Add 1 mL of 100 % Triton X-100 into 10 mL DEPC water (10 % v/v final). Store at room temperature.

   10× salt solution: 100 mM NaCl, 100 mM $MgCl_2$, 100 mM Tris–HCl pH 7.5 in DEPC water. Aliquot and store at −20 °C.

   1 M DTT: Add 1 g into 6.5 mL of DEPC water under the fume hood, and vortex well. Aliquot and store at −20 °C.

4. PBS-cycloheximide solution: Supplement 1× PBS with 10 μg/mL cycloheximide final using 10 mg/mL stock.
5. 1.5 mL microcentrifuge tubes.
6. Cell scrapers.
7. Microcentrifuge.
8. Filter barrier pipette tips.
9. Optional: DNase I, EDTA, puromycin.
10. Liquid nitrogen.

2.2 Sucrose Gradient Preparation and Fractionation

1. 1× gradient buffer: 30 mM Tris–HCl pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$. Store at 4 °C (*see* **Note 3**).
2. 10× gradient buffer: 0.3 M Tris–HCl pH 7.5, 1 M NaCl, 0.1 M $MgCl_2$. To prepare 1 L of 10× gradient buffer: 58.4 g NaCl, 20.34 g $MgCl_2$, 300 mL of 1 M Tris–HCl pH 7.5. Adjust the volume to 1 L with DEPC water. Store at 4 °C.
3. 50 % sucrose: 250 g sucrose in 500 mL of 1× gradient buffer, filter with 0.22 μm filter. Store at 4 °C.
4. 15 % sucrose: 75 g sucrose in 500 mL of 1× gradient buffer, filter with 0.22 μm. Store at 4 °C.
5. 3 % v/v hydrogen peroxide.
6. Rubber stoppers or parafilm.
7. Gradient former or a box lid/rack.
8. Polyallomer ultracentrifuge tubes (dimensions: 14 × 89 mm).
9. Ultracentrifuge.
10. Swinging bucket rotor.
11. Gradient analyzer (recommended: Teledyne Isco).
12. Tris Peristaltic Pump (recommended: Teledyne Isco).
13. UV spectrophotometer (UA-6 UV/VIS detector) (recommended: Teledyne Isco).
14. PeakTrack 110 program (recommended: Isco, Inc.).
15. Forceps.
16. 1.5 mL microcentrifuge tubes.

2.3 RNA Isolation from Sucrose Fractions

1. 20 mg/mL proteinase K.
2. 10 % w/v SDS in DEPC water.
3. 5 M NaCl in DEPC water.
4. Phenol:chloroform 5:1 acid equilibrated pH 4.7.
5. Isopropanol.
6. Water bath.
7. Forces and tissue paper.
8. RNase-free water.

2.4 Protein Isolation from Sucrose Fractions

1. 100 % Trichloroacetic acid (TCA).
2. Acetone: store at −20 °C.
3. 1× SDS-PAGE loading buffer: 60 mM Tris–HCl pH 8, 100 mM DTT, 2 % w/v SDS, 10 % v/v glycerol, 0.1 % bromophenol blue, ultrapure water. Aliquot and store at −20 °C.
4. Thermoblock.
5. Optional: Sonicator.

3 Methods

3.1 Preparation of Cell Lysates

1. Grow adherent cells in 100 mm dishes to ~80 % confluence (*see* **Note 4**).
2. Treat cells in culture with 100 μg/mL cycloheximide. Incubate at 37 °C for 3–4 min. Additional controls can be included by performing EDTA or puromycin treatment (*see* **Note 5**).

If possible, work in the cold room or on the bed of ice, using ice-cold solutions and keeping the cell dishes on a bed of ice all the time. Work quickly but without rushing.

3. Remove the medium using a vacuum pump. Let the plates drain on an angled bed of ice and aspirate the remaining medium. Wash each plate with 5 mL of ice-cold PBS-cycloheximide solution (*see* **Note 6**), removing carefully and completely the PBS-cycloheximide solution after each wash by aspiration (let the plates drain on an angled bed of ice) in order to avoid the dilution of the lysis buffer in the next step.
4. Add 300 μL of ice-cold lysis buffer directly to the dish, scrape, and transfer to a pre-chilled 1.5 mL tube (*see* **Note 7**).
5. Immediately place the samples on ice for 2 min, with occasional vortexing.
6. Centrifuge for 5 min at 16,000 × *g* at 4 °C to pellet the nuclei and cellular debris. Optional DNase I treatment can follow (*see* **Note 8**).
7. Freeze the lysates in liquid nitrogen and store at −80 °C or load directly on sucrose gradient by carefully layering it on the gradient surface (*see* **Note 9**).

3.2 Sucrose Gradient Preparation and Centrifugation

If possible, work in the cold room or on the bed of ice, keeping all solutions cold in order to preserve RNA. Keep rotor buckets cold throughout the procedure.

1. Use ultracentrifuge tubes extensively washed as follows (by filling the tubes completely): three washes with ultrapure water, three washes with DEPC water, 5-min incubation with 3 % v/v

H_2O_2 in DEPC water, and three washes with DEPC water. Then dry the tubes in the oven. For long-term storage remember to seal up the tubes (*see* **Note 10**).

2. Put the tubes on ice and add 5.5 mL of cold 50 % sucrose solution to each tube. Carefully overlay 15 % sucrose solution till the tube is completely filled, avoiding to disturb the interface. Close the tube with a rubber stopper or parafilm; no air bubbles must be trapped in the sucrose (*see* **Note 11**).
3. Gently plate the tubes in the sucrose gradient former at 4 °C. Program the device in the following manner: 10 min to lay down, 120 min in horizontal position, and 10 min to move back to vertical position. The gradients are now ready to be used (*see* **Note 12**).
4. If you use frozen samples, thaw lysates on ice 2 h before use (*see* **Note 13**).
5. Carefully remove 800 μL from the top of the sucrose gradient and gently (i.e., drop by drop, staying close to the surface) overlay the sucrose with 700 μL of cell lysate (*see* **Note 14**).
6. Using forceps if necessary, carefully lower the tubes into the buckets of the rotor and close them.
7. Ultracentrifuge the gradients at 274,000 × *g* for 1 h 40 min at 4 °C.
8. After the centrifugation, leave the tubes in their buckets for 10 min at 4 °C to allow the gradients to stabilize.

3.3 Sucrose Gradient Fractionation

1. Wash the tubing of the fractionator extensively with DEPC water, 3 % hydrogen peroxide, DEPC water again, and 50 % sucrose before starting to process the samples, making sure not to trap any air bubbles. Set the baseline with sucrose between 20 and 60.
2. Place a series of labeled 1.5 mL tubes in the fraction collector at the end of the flow cell.
3. Carefully remove the centrifuge tubes containing the sucrose gradients from the centrifuge rotor and mount them one by one on the collector device (in the meantime, keep the remaining tubes at 4 °C). Monitoring the absorbance at 254 nm, collect 1 mL fractions in the microcentrifuge tubes (Figs. 1a and 3a).
4. Keep the sucrose fractions at 4 °C when for immediate use; otherwise store them at −80 °C.

3.4 RNA Isolation from the Fractions

RNA can be isolated from individual fractions or the subpolysomal and polysomal fractions can be pooled (*see Fig. 3a*). *Scale the volumes below accordingly.*

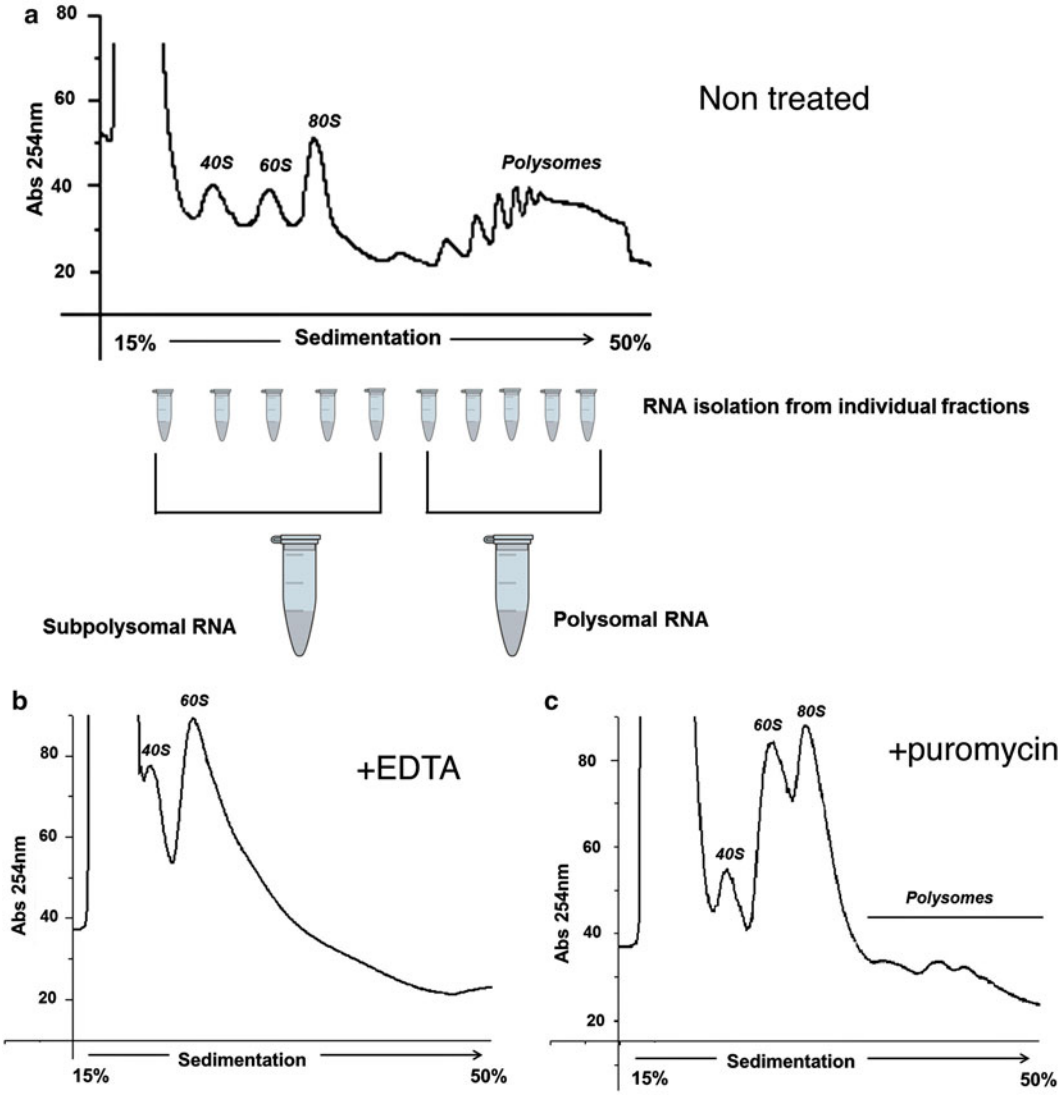

Fig. 3 (**a**) Representative polysome profile prepared from adherent cells in culture. 40S- and 60S-free small and large ribosomal subunits, 80S-monosomes. Pooling of fractions for downstream analysis is indicated. (**b**) Representative polysome profile after EDTA treatment: only 40S and 60S subunits can be seen. (**c**) Representative polysome profile after puromycin treatment: subunit and 80S peaks are increased and polysome peaks are decreased

1. To 1 mL fraction add 5 μL of proteinase K (to a final concentration of 100 μg/mL) and 100 μL SDS (to a final concentration of 1 %).
2. Digest the proteins by incubating the samples at 37 °C for 1–2 h in a water bath.
3. Add 250 μL of phenol:chloroform 5:1 acid equilibrated pH 4.7 to each 1 mL fraction and mix thoroughly by vortexing.

4. Add 100 μL of 5 M NaCl and vortex (*see* **Note 15**).
5. Centrifuge at 16,000 × *g* for 5 min at 4 °C.
6. Transfer the upper, aqueous phase to a fresh tube and add 1 mL of isopropanol (or 2 mL of ethanol). Mix and place at −80 °C for 1–2 h or overnight (recommended).
7. Centrifuge at 16,000 × *g* for 30–40 min.
8. Remove the supernatant and leave the pellet to air-dry (no more than 5 min). Resuspend the pellet in RNase-free water and proceed with purification using normal RNA precipitation or a commercial kit (*see* **Note 16**).

3.5 Protein Isolation from the Fractions

Work on ice under the fume hood.

1. To each 1 mL fraction add 100 μl of 100 % TCA and 1 mL of ice-cold acetone (*see* **Note 17**).
2. Put the sample at −80 °C overnight to induce protein precipitation.
3. Thaw the samples on ice and centrifuge at 16,000 × *g* for 5–10 min at 4 °C.
4. Remove the supernatant carefully leaving the white pellet intact (*see* **Note 18**).
5. Wash pellet with 1 mL of ice-cold acetone (*see* **Note 19**).
6. Centrifuge at 16,000 × *g* for 5 min at 4 °C. Remove the supernatant and repeat the acetone wash three times in total (*see* **Note 20**).
7. Dry the pellet by placing the tube under the fume hood for approximately 10 min to evaporate the acetone.
8. Resuspend the pellet in 50 μL of 1× SDS-PAGE loading buffer, vortex well, and incubate for 5–10 min at 100 °C (*see* **Notes 21** and **22**).

4 Notes

1. DEPC is toxic and irritant. Add DEPC to ultrapure water in 1 L bottle under the fume hood, close the cap tightly, and stir overnight using a magnetic stirrer. The following morning autoclave the bottles (remember to slightly loosen the caps) to deactivate DEPC.
2. Prepare 1× lysis buffer just before use because this solution is unstable.
3. Prepare 1× gradient buffer each time from 10× stock.

4. Once cells reach confluency, they decrease protein synthesis which results in lower polysome yield. Usually two 100 mm dishes are collected for each sample, but this is cell line dependent (for example, $1–2 \times 10^6$ cells for HeLa or $3–5 \times 10^6$ cells for HEK293). If you use a different cell line or special conditions (e.g., silencing, drug treatment) you might need to optimize the cell density of your culture/condition.
5. Cycloheximide inhibits protein synthesis and prevents ribosome disassembly; therefore, it is important to include it during all fractionation steps to preserve the polysome structures.

   To test whether transcripts are truly associated with polysomes it is possible to perform control experiments using EDTA or puromycin treatment. The intactness of ribosomes depends on Mg^{2+}. EDTA chelates Mg^{2+} and as a result dissociates the large and small ribosomal subunits and releases transcripts associated with ribosomes. In consequence only the first two peaks are seen, i.e., 40S and 60S (Fig. 3b). Puromycin inhibits protein translation and as a result polysomal peaks diminish and 80S increases (Fig. 3c). The sensitivity of the sedimentation of the transcripts to treatment with EDTA and puromycin suggests that they are associated with polysomes.

   Puromycin treatment: Add 100 μg/mL final puromycin directly to cells in a plate, and incubate for 15 min at 37 °C prior to lysate preparation.

   EDTA treatment: Add 50 mM EDTA final to cell lysate just before applying on gradient before ultracentrifugation. Incubate for 10 min on ice.
6. Add PBS on the side of the plate to avoid cell detachment.
7. If using two 100 mm dishes for one sample, add 150 μL to each of them and then pool.
8. Add DNase I to a final 0.005 U/μL and leave the lysates on ice for 30 min in order to allow the DNase I to degrade any DNA contamination.
9. The supernatant can be stored at −80 °C for a maximum of 6 months, but it is recommended to use the lysate within 1 month.
10. It is recommended to prepare the tubes the day before use.
11. The 50 % sucrose can be added using a pipette controller. The 15 % sucrose solution should be added drop by drop staying close to the interphase in order to preserve a sharp interface. Using 1000 μL pipette allows for a greater control. During the cap insertion or parafilm wrapping make sure that there are no air bubbles trapped inside the tube, since free-floating bubbles have a deleterious effect on gradient formation.

12. The gradients can be created manually. Transfer the tubes into a lid of sample storage box or a rack, and slowly (over 1 min) incline them by 90° so that the tubes are in a horizontal position. Keep them in this position for 2 h, and then again slowly move them back into the vertical position.
13. It is recommended to thaw lysates during the sucrose gradient formation to optimize the use of time.
14. Tubes can be reused; however, if they are reused too many times or are not completely filled, they will collapse during centrifugation. Therefore, the tubes need to be carefully controlled before and after each use. If the volume of cell lysate is less than 700 μL, adjust the volumes accordingly always leaving a space of 100 μL. If necessary, 1× gradient buffer can be added to cell lysate to obtain the required volume. Balance the tubes by applying the same volume of cell lysate.
15. Sodium acetate may be used as an alternative to NaCl when low-salt concentrations are needed in the final RNA sample.
16. After removing isopropanol, you can remove the remaining phenol traces from the tube with tissue paper and forceps. Any strong phenol contamination occurs because phenol slides from the tube walls down to the pellet; therefore, it is important to keep the tube upside down at all times. If you are skilful to remove the phenol with paper you can skip the cleanup step and resuspend RNA in RNase-free water directly.
17. TCA is extremely corrosive! Wear gloves and eye protection and work under the fume hood.
18. The pellet is not always visible, so it is important to remove the supernatant placing the tip on the opposite side of the tube where the pellet is expected to reside after centrifugation. The pellet will become visible in the next step.
19. Resuspend the pellet up and down many times, in order to remove any remaining traces of sucrose.
20. In these washes the pellets are white and compact.
21. Traces of TCA will turn the solution yellow. The use of 1× sample buffer at high pH should prevent this problem. If pH is not adjusted, the proteins might not run according to their molecular weight.
22. If the pellet is not dissolved after boiling, it might be necessary to sonicate it and then incubate at 100 °C for another 10 min.

    Immunoblotting is performed with the same volume of each sample in order to obtain the protein profile correspondent to the RNA profile; usually 25 μL is enough to visualize most proteins (Fig. 2b).

References

1. Warner JR, Knopf PM, Rich A (1963) A multiple ribosomal structure in protein synthesis. Proc Natl Acad Sci U S A 49:122–129
2. Kraushar ML, Thompson K, Wijeratne HRS et al (2014) Temporally defined neocortical translation and polysome assembly are determined by the RNA-binding protein Hu antigen R. Proc Natl Acad Sci U S A 111:E3815–E3824
3. Kuersten S, Radek A, Vogel C, Penalva LOF (2013) Translation regulation gets its "omics" moment. Wiley Interdiscip Rev RNA 4:617–630
4. Horvilleur E, Sbarrato T, Hill K et al (2014) A role for eukaryotic initiation factor 4B overexpression in the pathogenesis of diffuse large B-cell lymphoma. Leukemia 28:1092–1102
5. Arribere JA, Doudna JA, Gilbert WV (2011) Reconsidering movement of eukaryotic mRNAs between polysomes and P bodies. Mol Cell 44:745–758
6. Schwanhäusser B, Busse D, Li N et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–342
7. Arava Y, Wang Y, Storey JD et al (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 100:3889–3894
8. Tebaldi T, Re A, Viero G et al (2012) Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. BMC Genomics 13:220
9. Boussemart L, Malka-Mahieu H, Girault I et al (2014) eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies. Nature 513:105–109
10. Modelska A, Turro E, Russell R et al (2015) The malignant phenotype in breast cancer is driven by eIF4A1-mediated changes in the translational landscape. Cell Death Dis 6:E1603
11. Larsson O, Li S, Issaenko OA et al (2007) Eukaryotic translation initiation factor 4E induced progression of primary human mammary epithelial cells along the cancer pathway is associated with targeted translational deregulation of oncogenic drivers and inhibitors. Cancer Res 67:6814–6824

Chapter 5

Exploring Ribosome Positioning on Translating Transcripts with Ribosome Profiling

Pieter Spealman, Hao Wang, Gemma May, Carl Kingsford, and C. Joel McManus

Abstract

Recent technological advances (e.g., microarrays and massively parallel sequencing) have facilitated genome-wide measurement of many aspects of gene regulation. Ribosome profiling is a high-throughput sequencing method used to measure gene expression at the level of translation. This is accomplished by quantifying both the number of translating ribosomes and their locations on mRNA transcripts [1]. The inventors of this approach have published several methods papers detailing its implementation and addressing the basics of ribosome profiling data analysis [2–4]. Here we describe our lab's procedure, which differs in some respects from those published previously. In addition, we describe a data analysis pipeline, Ribomap, for ribosome profiling data. Ribomap allocates sequence reads to alternative mRNA isoforms, normalizes sequencing bias along transcripts using RNA-seq data, and outputs count vectors of per-codon ribosome occupancy for each transcript.

Key words Ribosome Profiling, Translation, Yeast, High-throughput sequencing, Bioinformatics, Ribomap, Ribo-seq

1 Introduction

Ribosome profiling simultaneously measures the relative number of ribosomes on a transcript (ribosome occupancy) and their locations along the transcript. Numerous studies have used ribosome occupancy to identify changes in translation based on environmental stresses [1, 5–7], developmental cues [8, 9], and divergence between species [10, 11]. Identifying the location of translating ribosomes has been instrumental in understanding mechanisms of translational regulation [12, 13], and has revealed that many RNA regions previously thought to be noncoding are actually engaged by ribosomes [1, 13, 14].

Ribosome profiling requires four steps to generate a sequencing library of ribosome protected fragments (RPFs): addition of a drug that inhibits translation, transcript digestion and purification,

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_5,

preparation of sequencing libraries from resulting RPFs, followed by sequencing and data analysis (Fig. 1). In order to estimate the relative efficiency of translation among mRNA transcripts, RNA-seq libraries must be prepared simultaneously from the same cell extracts. Many parameters can be varied, depending on the nature of the experiment. These options are briefly reviewed here before the "standard" method is described in detail.

Typically, chemical inhibitors of translation are used during ribosome profiling to preserve ribosome locations. Perhaps the most frequently used translation inhibitor, cycloheximide, binds along with tRNA to the E-site of the ribosome and prevents tRNA release. This causes the ribosome to stall on the transcript [15]. The ribosome is forced into a conformation that typically covers 27–30 nt of transcript, the standard footprint size of ribosome protected mRNA fragments [16].

Despite its prevalence of use, cycloheximide does have some limitations. Because cycloheximide inhibits elongation and not initiation, ribosomes can continue to initiate as long as the translation initiation site remains accessible [17]. At sufficient concentrations of cycloheximide, translation initiation sites are rapidly made inaccessible by cycloheximide-paused ribosomes and bias is minimized. Conversely, at lower concentrations ribosomes may translocate substantial distances down mRNA, allowing additional ribosomes access to the initiation site [17]. Such delayed inhibition by cycloheximide may produce concentration-dependent artifacts by biasing ribosome locations towards the 5′ end of *o*pen *r*eading *f*rames (ORFs) [15, 17]. While the standard concentration (used in this method) appears to avoid most of these artifacts, experiments sensitive to ribosome occupancy bias should consider using higher concentrations.

Unlike cycloheximide, harringtonine and lactimidomycin act specifically on pre-initiation ribosomes. Thus, treatment with either chemical allows post-initiation ribosomes to continue translating while capturing ribosomes at the location of initiation. Harringtonine and lactimidomycin have been used in ribosome profiling experiments to measure translation elongation rates and identify alternative translation initiation sites [18, 19].

The only alternative to chemical inhibition involves rapidly freezing cells with liquid nitrogen to halt translation [1]. Several studies that used cryogen-based inhibition reported significant differences in ribosome profiles compared to those using chemical treatments [18, 20]. These artifacts from chemical treatments may explain previously reported features such as translation elongation ramping [21] and excess signal at translation initiation sites [1]. However, the cryogenic approach may also result in inaccurate measurements of ribosome locations due to ribosome run-off during lysate preparation [1]. Furthermore, effective cryogenic preparations are difficult due to temperature fluctuations while performing ribosome profiling.

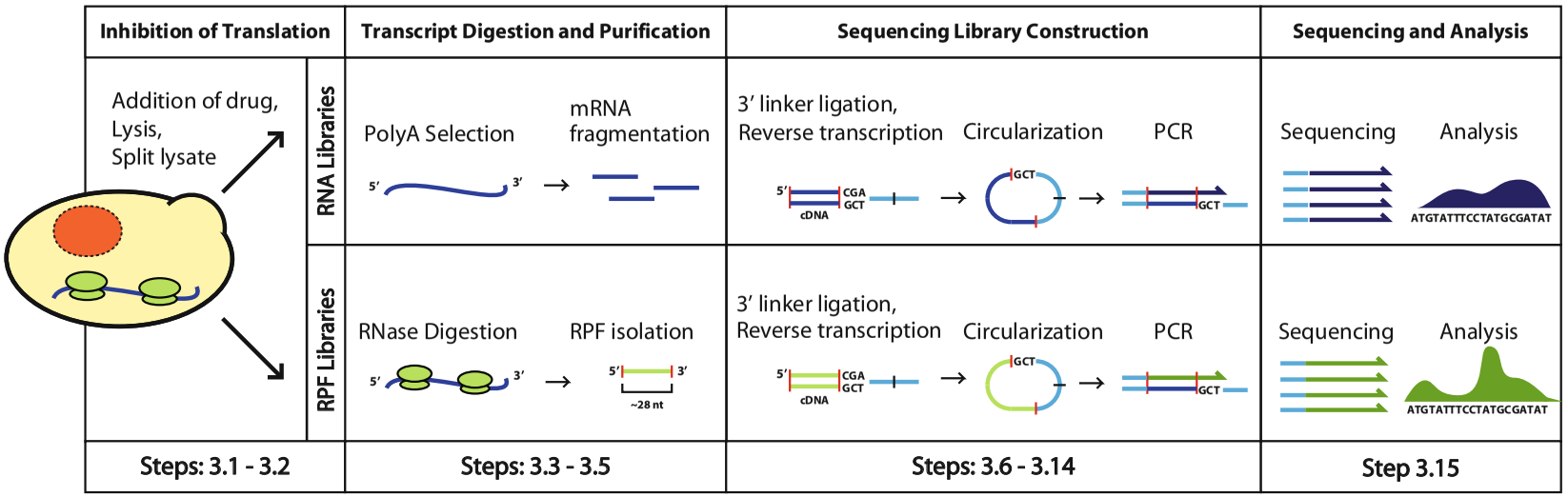

Fig. 1 Major steps in ribosome profiling. Ribosome profiling can be broken into four major phases. Subheadings 3.1 and 3.2 of this protocol describe the first phase, in which yeast cultures are grown to mid-log phase, treated with cycloheximide, flash-frozen, and lysed. The second phase of the procedure, detailed in Subheading 3.3 through 3.5 involves isolation of RNA fragments protected by the ribosome (RPFs) and from polyA enriched mRNA. Illumina sequencing library construction occurs in the third phase, and is described in Subheading 3.6 through 3.14. Finally, data analysis is outlined in Subheading 3.15 of this protocol

In its simplest form, ribosome profiling data can be analyzed to determine the fractional occupancy of ribosomes on each ORF (ribosome load). If RNA-seq data have been collected from matched samples, as described in this protocol, the fractional abundance of mRNA can also be calculated. Both ribosome loads and mRNA abundance are typically reported in a table as either read counts per gene, or as numbers normalized for the length of each ORF and the total number of reads in the sample (*r*eads *p*er *k*ilobase of gene model, per *m*illion mapped reads, or RPKM). The relative translation efficiency can be calculated as the quotient of the ribosome load and the mRNA abundance. It is important to note that these are compositional data, and thus should not be interpreted as the number of ribosomes on a gene or the number of mRNA molecules per cell. A few bioinformatic tools have been released that facilitate such gene level analyses, including ANOTA [22], and Babel [23].

In theory, ribosome profiling data provides nucleotide resolution measurements of ribosome position that could be used to calculate the per-codon occupancy of ribosomes on mRNA transcripts (ribosome profile vectors). However, there are several challenges to generating ribosome profile vectors in practice, including deconvolving multi-mapped reads, selecting the correct codon location that the P-site maps to, and correcting biases introduced during library preparation and sequencing. For example, one recently published approach designed to correct for sequencing bias [24] reports very different conclusions compared to prior studies that ignored it altogether. The final section of this protocol describes Ribomap [25], an automated pipeline we developed that simplifies ribosome profiling data analysis. Ribomap takes raw sequence reads from ribosome profiling and RNA-seq as input, and calculates ribosome loads, mRNA abundance, and translation efficiency for each transcript. Ribomap also accounts for multi-mapping sequence reads for both single and multiple isoform genes using RNA-seq estimates of isoform abundance.

2 Materials

2.1 Library Preparation Components

1. Micropipettors and filter tips.
2. Microcentrifuge.
3. 0.2 mL thin-walled tubes (certified nuclease free).
4. 1.5 mL tubes (nuclease free).
5. 50 mL conical tube.
6. 15 mL conical tubes.
7. 10 mL serological pipettes.
8. Stericup-GP, 0.22 μm, polyethersulfone, 500 mL (Millipore).
9. 0.5 mm diameter acid-washed glass beads.

10. Nuclease-free water.
11. 0.1 M Na_2CO_3.
12. 0.1 M $NaHCO_3$.
13. 3 M Sodium acetate (NaOAc) pH 5.5.
14. 1 M Sodium hydroxide (NaOH).
15. 1 M and 3 M Sodium chloride (NaCl) solutions prepared with nuclease-free water.
16. 20× SSC (Saline Sodium Citrate) (3 M NaCl, 0.3 M Sodium citrate, pH to 7.0 with 14 N HCl).
17. Tween 20.
18. 1 M Tris–HCl pH 7.4.
19. 10 mM dNTP mix.
20. 1 M and 100 mM DTT (Dithiothreitol).
21. 1 M Manganese chloride ($MnCl_2$).
22. 0.5 M EDTA pH 8.0.
23. Isopropanol.
24. Ethanol.
25. Sucrose.
26. DMSO (Dimethyl sulfoxide).
27. Acid phenol–chloroform pH 4.5 (Ambion).
28. Trizol.
29. Chloroform–isoamyl alcohol 24:1.
30. Antifoam A, molecular biology grade (Sigma).
31. 20 % SDS (Sodium lauryl sulfate).
32. Yeast tRNA (10 mg/mL).
33. 50 % PEG 8000.
34. GlycoBlue™ Coprecipitant (Life Technologies).
35. DNA Clean & Concentrator™—5 (Zymo Research).
36. Dynabeads® mRNA DIRECT™ Purification kit (Life Technologies).
37. T4 Polynucleotide Kinase (New England Biolabs).
38. T4 RNA Ligase 2, truncated (New England Biolabs).
39. Universal miRNA Cloning Linker (New England Biolabs).
40. Phusion® High-Fidelity DNA Polymerase (New England Biolabs).
41. RNAseI (100 U/μl) (Life Technologies).
42. SUPERase-In™ RNase Inhibitor (Life Technologies).
43. RNaseZAP® wipes (Life Technologies).
44. Dynabeads® MyOne™ Streptavidin C1 beads (Life Technologies).

45. SuperScript® II Reverse Transcriptase (Life Technologies).
46. CircLigase™ ss DNA Ligase (Epicentre).
47. Ultracentrifuge with TLA-100.3 rotor or equivalent.
48. Ultracentrifuge tubes (Beckman Coulter polyallomer #326823).
49. Thermocycler.
50. Thermomixer (or heat block).
51. Tube rotator.
52. DynaMag™—2 Magnet (Life Technologies).
53. Tapestation, D1000 Tapes and reagents or Bioanalyzer (Agilent) (optional).
54. NanoDrop.
55. Vortexer.
56. Vacuum pump (optional, we use Welch Model 2522B-01).
57. Custom Oligos (*see* Table 1).
58. Polysome Lysis Buffer: 20 mM Tris–HCl pH 8, 140 mM KCl, 1.5 mM $MgCl_2$, 1 % Triton X-100, 100 μg/mL cycloheximide (*see* Table 2).
59. Sucrose cushion solution: 1 M sucrose, 20 mM Tris–HCl pH 7.4, 250 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 0.1 U/μl SUPERase-in (*see* Table 3).

Table 1
Custom oligos (*see* Note 15)

RNA marker oligo 26 nt	5′AUGUACACGGAGACCCGCAACGCGA3′[Phos]
RNA marker oligo 34 nt	5′AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA3′ [Phos]
Reverse transcription (RT) primer	5′(Phos)AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT AGATCTCGGTGGTCGC(SpC18)CACTCA(SpC18) TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCT ACAG-3′
Subtractive hybridization primer 1	5′(BioTEG)AAGAGGTGCACAATCGACCGATCCTGA3′
Subtractive hybridization primer 2	5′(Biosg)TAGTTTCTTTACTTATTCAATGAAGCGG3′
Subtractive hybridization primer 3	5′(Biosg)AATATAGATGGATACGAATAAGGCGTC3′
Subtractive hybridization primer 4	5′(Biosg)TGGCTTAGTGAGGCCTCAGGATCTGCT3′
Subtractive hybridization primer 5	5′(Biosg)TCGAAGAGTCGAGTTGTTTGGGAATGC3′
Subtractive hybridization primer 6	5′(Biosg)CTATGCCGACTAGGGATCGGGTGGTG3′
Universal Forward PCR primer	5′-AATGATACGGCGACCACCGAGATCTACAC-3′
Barcoded Reverse PCR primer (*see* **Note 15**)	5′-CAAGCAGAAGACGGCATACGAGAT-(6 nt Illumina barcode)-TGTACTGGAGTTCAGACGTGTGCTCTTCCG-3′

Table 2
Polysome lysis buffer (make fresh)

Stock	Amount	Final concentration (in 10 mL)
Ultrapure (or DEPC) H_2O	8 mL	–
1 M Tris–HCl pH 8.0	200 µl	20 mM
1 M KCl	1.4 mL	140 mM
1 M $MgCl_2$	15 µl	1.5 mM
Triton X-100	100 µl	1 %
Cycloheximide (fresh stock)	100 µl (50 mg/mL)	100 µg/mL

Table 3
Sucrose cushion solution (add SUPERase-in immediately before use)

Reagent	Amount per sample	Final concentration
Sucrose	1.15 g	1 M
Nuclease-free water	2.36 mL (fill to 3.33 mL)	N/A
1 M Tris–HCl pH 7.4	66 µl	20 mM
1 M NaCl	830 µl	250 mM
1 M $MgCl_2$	50 µl	15 mM
1 M DTT	3.3 µl	1 mM
SUPERase-in™ (20 U/µl)	16.6 µl	0.1 U/µl

60. 2× Alkaline fragmentation buffer: 2 mM EDTA, 0.1 M 0.1 M Na_2CO_3, 0.1 M, 0.1 M $NaHCO_3$, pH to 9.2 (*see* Table 4).
61. YEPD medium: 1 % yeast extract, 2 % peptone, 2 % dextrose (*see* Table 5).
62. PAGE RNA extraction buffer: 300 mM NaOAc, pH 5.5, 1 mM EDTA (*see* Table 6).
63. PAGE DNA extraction buffer: 300 mM NaCl, 10 mM Tris–HCl pH 8, 1 mM EDTA (*see* Table 7).
64. Subtractive hybridization buffer: 2 M NaCl, 1 mM EDTA pH 8, 10 mM Tris pH 7.4, 0.02 % Tween 20 (*see* Table 8).

2.2 PAGE Components

1. Polyacrylamide gel electrophoresis equipment—cooled vertical unit with 18 × 16 cm glass plates, 15 well combs (1.5 mm thick), and 1.5 mm thick spacers.
2. Light box/Dark reader camera.

Table 4
2× Alkaline fragmentation buffer (pH to 9.2 with NaOH and HCl; store at –20 °C)

Reagent	Amount	Final concentration
0.5 M EDTA pH 8.0	80 µl	2 mM
0.1 M Na_2CO_3	2.4 mL	0.1 M
0.1 M $NaHCO_3$	17.52 mL	0.1 M

Table 5
YEPD (1 L)

Reagent	Amount	Final concentration
(A) YEP solution		
Peptone	20 g	2 %
Yeast extract	10 g	1 %
Deionized water	Fill to 900 mL	NA
(B) 20 % Dextrose		
Dextrose	20 g	20 %
Deionized water	Fill to 100 mL	NA

Autoclave YEP and 20 % Dextrose separately and combine before use to make 1 L YEPD

Table 6
PAGE RNA extraction buffer

Reagent	Amount	Final concentration
Nuclease-free water	22.33 mL	NA
3 M NaOAc (pH 5.5)	2.5 mL	300 mM
0.5 M EDTA	50 µl	1 mM

Table 7
PAGE DNA extraction buffer

Reagent	Amount	Final concentration
Nuclease-free water	6.88 mL	NA
1 M NaCl	3 mL	300 mM
1 M Tris pH 8	100 µl	10 mM
0.5 M EDTA	20 µl	1 mM

Table 8
Subtractive hybridization buffer

Reagent	Volume	Final concentration
3 M NaCl	3.33 mL	2 M
0.5 M EDTA pH 8	10 µl	1 mM
1 M Tris pH 7.4	50 µl	10 mM
Nuclease-free water	1.6 mL	–
Tween 20	1 µl	0.02 %

3. Dark reader.
4. Plastic wrap.
5. Corning® Costar® Spin-X® centrifuge tube filters, cellulose acetate membrane, 0.22 µm pore.
6. EMD Millipore Steriflip™ sterile disposable vacuum filter unit 0.22 µm PVDF.
7. 50 mL syringe.
8. 18G × 1½ needles.
9. Razor blades.
10. Vacuum source.
11. 10× TBE.
12. Urea.
13. 40 % Acrylamide–Bis solution 19:1.
14. TEMED (Tetramethylethylenediamine).
15. Ammonium persulfate ($(NH_4))2S_2O_8$).
16. 2× formamide loading buffer (95 % deionized formamide, 0.5 mM EDTA pH 8, 0.005 g bromophenol blue per 10 mL).
17. 10 bp DNA ladder (Life Technologies).
18. SYBR® Gold Nucleic Acid Gel Stain, 10,000× (Life Technologies).

3 Methods

3.1 Yeast Culture Preparation

3.1.1 Prepare Yeast Culture

The following describes a procedure to rapidly process yeast samples within minutes of the addition of cycloheximide.

1. Start a 100 mL overnight culture the night before you wish to harvest the cells.
2. First thing in the morning, restart yeast cultures at OD_{600} ~0.2 in 650 mL of YEPD in 2.8 L Fernbach Culture flasks from Corning (*see* **Note 1**).

3.1.2 Prepare Polysome Lysis Buffer

1. Make a fresh stock of cycloheximide (150 mg cycloheximide in 3 mL EtOH for a final concentration of 50 mg/mL). Cycloheximide is highly toxic. Stocks should be prepared in a fume hood.
2. Make 10 mL polysome lysis buffer (PLB, Table 2):
3. Filter-sterilize PLB using a 50 mL conical Steriflip with a 0.22 μm filter and a vacuum.
4. For each culture, fill a 50 mL conical tube with 2.2 mL of PLB. Place these in an ice water slurry for use at **step 6** in Subheading 3.1.3.
5. For each culture, place one 10 mL serological pipette to the side.

3.1.3 Prepare Flash Freeze Setup

1. Get roughly 1 L of liquid nitrogen, $N_2(l)$ for each culture in a container rated for $N_2(l)$.
2. Select styrofoam shipping containers large enough to hold one 50 mL conical tube per sample (*see* **Note 2**).
3. Using a 16 gauge needle, carefully punch 4 holes in the lids of 50 mL conical tubes (one per culture). Label the conical tubes and lids with the sample names using a permanent marker. Place the lids to the side.
4. Place the 50 mL conical tubes in a 3×3 rack in the styrofoam container selected in **step 2**.
5. Fill each conical tube three quarters full with $N_2(l)$. Pour additional $N_2(l)$ into the container to a depth of 3–4 cm.

3.1.4 Cell Harvesting and Collection by Vacuum Filtration

The final cycloheximide concentration should be at least 100 μg/mL. Be sure to follow standard chemical safety procedures when handling cycloheximide in both liquid and powder forms.

1. Assemble the 500 mL Stericup-GP vacuum filters per manufacturer's specifications (*see* **Note 3**).
2. Add 1.3 mL of 50 mg/mL cycloheximide/ethanol stock to one culture (100 μg/mL final concentration). Replace culture flask in shaker for an additional 2 min (*see* **Note 4**).
3. Ensure that the 10 mL serological pipettes, the ice slurry, and 50 mL conical tubes containing PLB are within reach. Ensure that the container being used for flash freezing is nearby and still has ~10 mL of $N_2(l)$ left in each tube.
4. After 2 min has elapsed since **step 2**, add 50 μl of Sigma Antifoam A (this step is optional, but greatly helps with filter sterilization if yeast culture is foamy) and proceed immediately to **step 5**.
5. Harvest cells by pouring the contents into the Stericup filtration system. Ensure that the cells remain suspended by gently

swirling the growth flask until it has been entirely emptied. CAUTION—the media filtrate contains cycloheximide.

6. Once the media has been filtered, carefully scrape the cell paste off the filter using a flat edged spatula. Spoon all of the cell paste into a conical tube of ice cold PLB from **step 4**, Subheading 3.1.1.
7. Use a 10 mL serological pipette to thoroughly resuspend the cells in PLB. We use an electronic pipettor to assist with resuspension.
8. Once the cells are homogenously mixed with PLB, flash-freeze the cells by slowly dripping the cell/PLB suspension into the labeled 50 mL conical tube containing $N_2(l)$. Add $N_2(l)$ as needed to ensure frozen cell droplets remain submerged during this process.
9. Close each tube with the correctly labeled cap.
10. Before moving to process the next sample, ensure that 3–4 cm of $N_2(l)$ remains in the bottom of the styrofoam container.
11. Repeat **steps 1–9** for any additional cultures.
12. When each sample has been processed the tube rack may be remove from the container and placed in a −80 °C freezer to allow for the $N_2(l)$ to evaporate.

NOTE: Stopping point. Frozen cells can be stored at −80 °C for months or used immediately for lysis.

3.2 Yeast Cell Lysis

Yeast require significant mechanical disruption to lyse their cell wall. This is further complicated in ribosome profiling, as lysis must occur at low temperatures. The most conservative solution uses specialized equipment (e.g., a Retsch cryo-mill) that pulverizes cells under cryogenic conditions [1]. However, this may be unnecessary when sufficient levels of translation inhibitors are used. We describe an alternative method that lyses cells by alternating cycles of vortexing with glass beads and immersion in ice water, such that the temperature is maintained within 3 ° C of freezing at all times. Our lab has found minimal differences in ribosome profiles we generated with either method. However, we believe that cryogenic lysis is necessary for experiments done in the absence of translation inhibitors.

1. Fill a $N_2(l)$ dewar with ~0.5 L of $N_2(l)$, in preparation for **step 15.**
2. Thaw the flash-frozen cells from Subheading 3.1.3 **step 12** in an ice water slurry.
3. For each sample, label and fill a 15 mL conical tube with 2 mL of 0.5 mm diameter acid-washed glass beads.
4. Resuspend the thawed cell/PLB suspension with a 10 mL pipette and transfer to the labeled 15 mL conical tube. Ensure the caps are on tight.

5. Return tubes to the ice water slurry for 5 min.
6. Vortex each sample at max for 5 s, fully immerse in ice slurry for 30 s.
7. Repeat **step 6**, 11 times for each sample (for a total of 12 rounds of vortexing).
8. Separate the lysate from the glass beads by centrifuging for 5 min at 3000 × *g* at 4 °C.
9. Transfer supernatants to clean, labeled, 1.5 mL microcentrifuge tubes.
10. Clarify lysate by centrifuging for 10 min at 20,000 × *g* at 4 °C.
11. Recover the supernatant, taking care to avoid the pellet and lipid layer at the surface. Repeat **step 9** if the recovered sample is visibly contaminated with pellet or lipid material.
12. Make a 1:200 dilution in a microcentrifuge tube (5 μl extract with 995 μl sterile deionized H_2O) for each sample.
13. Using the NanoDrop measure the A_{260} of each diluted sample and calculate the OD_{260}/mL.
14. Using PLB (with fresh cycloheximide if there has been greater than 1 week since its creation) dilute the sample to 200 OD_{260}/mL.
15. Separate this into 250 μl aliquots (50 OD_{260}) per tube.
16. Flash-freeze these using the $N_2(l)$ from **step 1** and store at −80 °C.

3.3 Preparing Ribosome Protected Fragments (RPFs)

RNase I is added to cell lysate to digest mRNA regions not bound by ribosomes. This results in a heterogeneous mixture of partially digested mRNA, tRNA, ribonucleoprotein complexes, and single ribosomes with their protected fragment (also referred to as monosomes). RNase I is effective for yeast, however micrococcal nuclease has been used to digest lysates from other species [12, 13]. After nuclease digestion, monosomes must be separated from other RNA species. This is accomplished by ultracentrifugation through either a sucrose gradient [1] or a sucrose cushion [18]. Sucrose cushions effectively purify monosomes, are easy to construct, and are thus suggested for most experiments. However, sucrose gradient fractionation may allow more precise selection of monosomes.

3.3.1 Nuclease Digestion and Ultracentrifugation

1. Thaw one aliquot of lysate (50 ODs) per sample and bring the volume to 350 μl with polysome lysis buffer (*see* Subheading 3.1.1 **step 2**).
2. Add 8 μl RNaseI (100 U/μl) to each sample.
3. Incubate the lysate for 50 min at room temperature on a tube rotator.
4. During the RNAseI digestion, prepare 3.3 mL of sucrose cushion solution per sample (Table 3).

5. Filter sterilize sucrose cushion solution using a 50 mL EMD Millipore Steriflip™ sterile disposable vacuum filter unit (0.22 μm PVDF) and place the solution on ice.
6. Clean the ultracentrifuge tubes with RNAseZap® wipes and rinse twice with nuclease-free water.
7. After 50 min, stop the RNAseI digests by adding 10 μl SUPERase-in™ to each tube and place them on ice.
8. Aliquot 3 mL of sucrose cushion solution to each ultracentrifuge tube.
9. Slowly pipette each extract onto the top of a sucrose cushion.
10. Balance the centrifuge tubes to within 0.05 g of each other before loading them into the rotor (*see* **Note 5**).
11. Load the sucrose cushions in a TLA-100.3 rotor (or equivalent) and spin 4 h at 70,000×*g* (*see* **Note 6**).

3.3.2 RNA Extraction

1. Carefully remove the centrifuge rotor and immediately pipette off sucrose supernatants, leaving the pellets behind. The pellets look like clear lenses in bottom rear of tube.
2. Resuspend each pellet in 712 μl 10 mM Tris, pH 7.4 and transfer to a 2 mL microcentrifuge tube.
3. Add 38 μl of 20 % SDS
4. Add 1 volume acid phenol–chloroform (pH 4.5).
5. Incubate the samples at 65 °C thermomixer for 5 min with maximum shaking (*see* **Note 7**).
6. Place the tubes on ice for 5 min.
7. Spin at full speed on tabletop centrifuge for 5 min.
8. Separate aqueous phase (top phase) into new microcentrifuge tube. Take care to not pipette up any of the white interphase layer.
9. Repeat **steps 4–8** once (for a total of two rounds).
10. Add 1 volume of chloroform–isoamyl alcohol 24:1 and vortex the sample for 1 min at room temperature.
11. Spin the samples in a microcentrifuge at top speed for 5 min at room temperature.
12. Remove aqueous phase (top) into new microcentrifuge tube.
13. Add one-tenth volume of 3 M NaOAc (pH 5.5) to each tube.
14. Add 1 volume of isopropanol.
15. Precipitate the RNA at either −80 °C for 30 min, or −20 °C for 1 h. Alternatively, the protocol can be stopped here and the tubes can be left at −20 °C overnight.
16. To pellet the RNA, centrifuge the precipitated RNA in at the maximum speed at 4 °C in a microcentrifuge for 30 min.

17. Carefully pipette off the supernatant, leaving the RNA pellet behind.
18. Wash each RNA pellet with 500 µl 70 % ethanol until it is freely floating. Take care to ensure the pellet does not stick to the pipette tip.
19. Centrifuge samples at maximum speed in a 4 °C microcentrifuge for 5 min.
20. Pipette off the ethanol supernatant. We typically remove 450 µl with a P1000, quickly spin down the samples, and remove the remainder with a P200. Open the caps to allow the pellets to dry for 1–2 min.
21. Pipette up and down to thoroughly resuspend each RNA pellet in 20 µl of nuclease-free water.
22. Measure the concentration of the RNA with a NanoDrop. These RPF samples can be stored at −20 °C.

3.4 Preparing mRNA Fragments

To estimate the relative translation efficiency of transcripts (ribosomes per transcript), mRNA abundance must be measured in parallel using RNA-seq. Because ribosomal RNA (rRNA) comprises more than 95 % of total cellular RNA, measures are taken to enrich mRNA and deplete rRNA (*see* **Note 8**).

3.4.1 Total RNA Extraction

1. Bring 10 OD 260 units of cell lysate (Subheading 3.2 **step 16**) to 712 µl by adding 10 mM Tris, pH 7.4 and transfer to a 2 mL microcentrifuge tube.
2. Extract the RNA (Subheading 3.3.2 **steps 3** through 21).
3. Measure the RNA concentration with a NanoDrop. These samples should be used directly for mRNA enrichment, as freeze–thaw cycles may lead to mRNA fragmentation.

3.4.2 mRNA Enrichment with the Dynabeads® mRNA DIRECT™ Purification Kit

1. Vortex the Dynabeads® Oligo(dT) beads and pipette 250 µl into one 1.5 mL microcentrifuge tube per sample.
2. Place tubes on a DynaMag™—2 Magnet for 1 min. Pipette off the supernatant.
3. Remove the tubes with the beads from the magnet and resuspend the beads in 125 µl lysis/binding buffer (provided with the kit) and place on the magnet for 1 min.
4. Pipette off liquid and resuspend beads in 125 µl lysis/binding buffer.
5. Place bead tubes on magnet.
6. Dilute 100 µg of total RNA to a final volume of 125 µl using nuclease-free water.
7. Heat the RNA to 65 °C for 2 min.

8. Place the tubes on ice and add 1 μl SUPERase-in™.
9. Add the RNA to the beads in lysis/binding buffer.
10. Incubate the mixture at room temperature for 5 min in a tube rotator.
11. Place the tubes on the magnet for 1 min and remove the unbound liquid.
12. Resuspend the beads in 200 μl wash buffer B (provided with the kit), vortex gently to mix, and place back on the magnet. Remove the wash buffer and repeat this step (total of two washes).
13. Resuspend the beads in 125 μl 10 mM Tris elution buffer (provided with the kit).
14. Heat RNA at 80 °C for 2 min to elute mRNA from beads.
15. Immediately place the sample on the magnet. Incubate for 30 s and pipette off the RNA sample into a fresh 1.5 mL tube containing 125 μl of lysis/binding buffer.
16. Place tubes on ice while performing **step 17**.
17. Wash the previously used beads twice with 200 μl wash buffer B (*see* **step 12**).
18. Add the RNA to the beads (be sure to keep the same sample on the same previously used beads).
19. Repeat **steps 10** through **12**.
20. Resuspend the beads in 20 μl 10 mM Tris elution buffer.
21. Heat RNA by to 80 °C for 2 min to elute mRNA from beads.
22. Immediately place the sample on the magnet. Incubate for 30 s and pipette off the RNA sample into a 0.2 mL thin-walled tube.

3.4.3 mRNA Fragmentation

1. In a 0.2 mL thin-walled tube, mix 20 μl mRNA with 20 μl of 2× alkaline fragmentation buffer (Table 4).
2. Incubate the samples for 40 min at 95 °C in a thermocycler with a heated lid, and place the reaction on ice for 5 min.
3. Spin tubes down and transfer to labeled 1.5 mL tubes.
4. Add 60 μl 3 M NaOAc pH 5.5, 2 μl GlycoBlue™, and 500 μl nuclease-free water.
5. Add 600 μl isopropanol to each tube.
6. Precipitate the RNA as described in Subheading 3.3.2, **steps 15–20**.
7. Pipette up and down to thoroughly resuspend each RNA pellet in 20 μl of nuclease-free water.
8. Using a NanoDrop, measure the concentration and quality of the RNA. These fragmented mRNA samples can be stored at −20 °C.

3.5 RNA Fragment Size Selection Using Polyacrylamide Gel Electrophoresis

At this point in the procedure, nuclease treated RPF samples are almost entirely comprised of rRNA, as the mRNA protected by each ribosome is ~0.5 % the length of the ribosome. Polyacrylamide Gel Electrophoresis (PAGE) is used to separate the ribosome protected mRNA fragments from most of the rRNA. The procedure below effectively purifies mRNA fragments ~27–32 nucleotides in length, the typical ribosome footprinting size in cycloheximide treated cells (*see* **Note 9**).

1. In a 50 mL conical tube, prepare 40 mL of 15 % polyacrylamide 8 M Urea gel (*see* **Note 10**).
2. Once the urea has dissolved, filter the gel using a 50 mL 0.22 μm Steriflip™. If the solution is warm, place it in a beaker with room temperature water to cool.
3. Set up the vertical gel apparatus, glass plates, and 1.5 mM spacers so that it is ready to be poured.
4. Add 200 μl of 10 % ammonium persulfate and 20 μl of TEMED to the 40 mL of filtered 15 % polyacrylamide and urea. Mix well by gentle inversion and immediately pour the gel.
5. Insert the gel comb and allow the gel to polymerize at least 30 min at room temperature (*see* **Note 11**).
6. Add 10 μl 2× formamide load dye to each sample.
7. Prepare the 10 bp DNA ladder by adding 1 μl ladder to 9 μl nuclease-free water and 10 μl 2× formamide load dye.
8. Prepare RNA marker oligos (26 and 34 nt) by resuspending to a 10 μM stock in nuclease-free water. Then, add 10 pmol of each oligo (1 μl of each 10 μM stock) to 3 μl nuclease-free water. Add 5 μl 2× formamide load dye.
9. Heat all the samples, ladders, and RNA marker oligos, at 80 °C for 3 min and place on ice.
10. Once the gel has polymerized, place the gel into the gel box and pour 1× TBE into the top and bottom reservoirs so that the electrodes are submerged.
11. Fill a syringe with a needle attached with 1× TBE and blow the urea out from each well.
12. Pre-run the gel at 300 V for at least 20 min. For different sized gels, adjust the voltage to maintain an equivalent V/cm.
13. After pre-running the gel, fill a 10 mL syringe with a needle attached with 1× TBE and rinse each gel well to remove urea that leaches out during the pre-run.
14. Immediately load the denatured RNA onto the gel. Load the RNA marker oligos next to the RPF samples every few lanes (approximately four). For the mRNA samples load the 10 bp ladder every four lanes to ensure accurate sizing.

15. Run the gel at 300 V until the bromophenol blue is near the bottom. The bromophenol blue runs at about 10 nt. This takes approximately 2 h.
16. Prepare a spin extractor for each sample by puncturing a hole at the bottom of a 0.5 mL tube. Place the 0.5 mL tube inside of a 2 mL tube.
17. Remove the gel from the apparatus and place it in a container lined with plastic wrap. Add 100 mL of 1× TBE and 10 μl of SYBR® Gold nucleic acid stain and mix briefly. Incubate for 5 min at room temperature.
18. Visualize the gel on a dark reader (preferable) or a UV illuminator.
19. For the RPF samples, excise the region between the two RNA marker oligos (from 26 to 33 nt) using razor blades. For the mRNA samples, excise the region between 30 and 50 nt. Use a new razor blade to cut out each sample. The marker oligos can be excised and used as a positive control for later library preparation steps.
20. Place cut out gel bands in the top tube of spin extractors.
21. Spin the spin extractors for 2 min in a centrifuge at top speed to fragment gel into the 2 mL tube.
22. Add 400 μl of PAGE RNA extraction buffer (Table 6) to the 2 mL tube containing the fragmented gel. Add 2 μl of SUPERase-In™ to each tube.
23. Incubate at room temperature on a tube rotator or mixer overnight.
24. For each sample, add the gel slurry and liquid to a Costar® Spin-X® centrifuge tube filter.
25. In a microcentrifuge, spin tubes for 2 min at top speed.
26. Discard the filter containing the gel slurry.
27. Add 1 μl GlycoBlue™ to each tube.
28. Add 1 volume of isopropanol.
29. Precipitate the RNA as described in Subheading 3.3.2, **steps 15–20**.
30. Pipette up and down to thoroughly resuspend each RNA pellet in 11.5 μl nuclease-free water.

3.6 Repair RNA 3′ Ends

Both nuclease digestion (RPF) and base hydrolysis (mRNA) can leave cyclic 2′–3′ phosphates on the 3′ end of RNA fragments. This step repairs these cyclic ends to generate clean 3′ OH for later library preparation steps.

1. In a 0.2 mL thin-walled tube, heat the RNA sample at 80 °C for 2 min and place on ice.

2. Add the following to each RNA sample (total volume 15 μl):

Reagent	Per reaction
10× T4 PNK buffer	1.5 μl
SUPERase-In® (20 U/μl)	1 μl
T4 PNK	1 μl

3. Incubate at 37 °C for 1 h, then heat inactivate by incubating the samples at 70 °C for 10 min.
4. To each sample, add 39 μl of nuclease-free water, 10 μl of 3 M NaOAc (pH 5.5), 1 μl of GlycoBlue™, and 150 μl of isopropanol.
5. Precipitate the RNA as described in Subheading 3.3.2, **steps 15**–20.
6. Pipette up and down to thoroughly resuspend each RNA pellet in 9 μl of nuclease-free water.

3.7 Ligate the 3′ Linker

1. Transfer each sample to a 0.2 mL thin-walled tube.
2. Add 1 μl (250 ng/μl) NEB miRNA cloning linker to each sample.
3. In a thermocycler, heat the RNA to 80 °C for 2 min. Place the samples on ice.
4. Add 10 μl of reaction mix (below) to each sample:

Reagent	Per reaction
50 % PEG 8000	5 μl
DMSO	2 μl
10× RNA ligase buffer	2 μl
SUPERase-In® (20 U/μl)	1.0 μl

5. Add 1 μl T4 RNA ligase 2, truncated, to each sample.
6. Incubate for 2.5 h at 22 °C. Alternatively, incubate overnight at 16 °C.
7. Add 338 μl of nuclease-free water, 40 μl of 3 M NaOAc (pH 5.5), 1.5 μl of GlycoBlue™, and 500 μl of isopropanol to each sample.
8. Precipitate the RNA as described in Subheading 3.3.2, **steps 15–20**.
9. Pipette up and down to thoroughly resuspend each RNA pellet in 10 μl of nuclease-free water.

3.8 PAGE Purification of Ligation Products

1. Refer to Subheading 3.5 Size selection using Polyacrylamide Gel Electrophoresis and ethanol precipitation for this procedure, with the following modifications.
2. Prepare a 12 % Polyacrylamide 8 M Urea gel (**steps 1–5** of Subheading 3.5).
3. Add 10 μl of 2× formamide loading dye to each sample.
4. Prepare 10 bp ladder by adding 1 μl ladder to 9 μl nuclease-free water and 10 μl 2× formamide load dye.
5. Run samples on the gel as described in **steps 9–18** of Subheading 3.5.
6. For the RPF samples, excise the region between 43 and 51 nt. If the dephosphorylation and ligation reactions were successful, then the marker oligos should now be at 43 and 51 nt.
7. For the mRNA samples, excise the region between 50 and 70 nt.
8. Prepare gel elutions as described in **steps 20–23** of Subheading 3.5.
9. Follow **steps 24–29** of Subheading 3.5 to precipitate RNA from gel eluates.
10. Resuspend each RNA pellet in 10 μl of nuclease-free water.

3.9 Reverse Transcription

1. Add 2 μl of 2.5 μM RT primer to 10 μl of each RNA sample.
2. On a thermocycler, heat the RNA and primer mix by incubating for 2 min at 80 °C followed by 5 min at 65 °C. Place the RNA and primer mix on ice.
3. Add 7 μl reaction mix (below):

Reagent	Per Reaction
5× First strand buffer	4 μl
10 mM dNTPs	1 μl
100 mM DTT	1 μl
SUPERase-In® (20 U/μl)	1 μl

4. Add 1 μl Superscript® III to each reaction.
5. On a thermocycler, incubate at 48 °C for 30 min.
6. Destroy RNA template by adding 2.0 μl of 1 M NaOH and incubating for 20 min at 98 °C on a thermocycler. Place on ice.
7. Add 156.5 μl nuclease-free water, 20 μl of 3 M NaOAc pH 5.5, and 1.5 μl of glycoblue to each sample.
8. Add 300 μl isopropanol to each sample and precipitate cDNA by incubating −80 °C for 30 min. Alternatively, precipitate at −20 °C for 1 h or overnight.

9. To pellet the cDNA, centrifuge for 30 min at 4 °C at 20,000 × *g*.
10. Remove and discard supernatant (*see* **Note 12**).
11. Wash the pellet by adding 600 µl of 70 % ethanol. Centrifuge the samples for 15 min at 4 °C at 20,000 × *g*.
12. Remove all of the 70 % ethanol and resuspend each cDNA pellet in 10 µl nuclease-free water. The sample can be stored at −20 °C. Alternatively, proceed directly to the next step.

3.10 PAGE Purification of cDNA Products

1. Refer to Subheading 3.5, for this procedure, with the following modifications:
2. Prepare a 10 % Polyacrylamide 8 M Urea gel (*see* **steps 1–5** in Subheading 3.5).
3. Add 10 µl of 2× formamide loading dye to each sample.
4. Prepare 10 bp ladder by adding 1 µl ladder to 9 µl nuclease-free water and 10 µl 2× formamide load dye.
5. For RPF samples, the extended products should be around 130 bp, while unextended primer should run at around 100 nt. mRNA samples products should be ~140–160 nt long. Excise the product bands using a new clean razor blade for each sample.
6. Prepare gel elutions as described in **steps 20–23** of Subheading 3.5, using the DNA elution buffer (Table 7).
7. Follow **steps 24–29** of Subheading 3.5 to precipitate cDNA from gel eluates.
8. Resuspend each cDNA pellet in 15 µl of nuclease-free water.

3.11 Circularization

1. Transfer 7.5 µl of each sample into a 0.2 mL thin-walled tube. The remaining 7.5 µl of sample can be stored at −20 °C.
2. Add 2 µl Reaction mix (below) to each sample and mix well.

Reagent	Per reaction
10× CircLigase Buffer	1 µl
1 mM ATP	0.5 µl
50 mM $MnCl_2$	0.5 µl

3. Add 0.5 µl CircLigase to each sample.
4. In a thermocycler, incubate for 1.5 h at 60 °C.
5. Heat inactivate by incubating the sample at 80 °C for 10 min. The sample can be stored at −20 °C. Alternatively, proceed to the next step.

3.12 Subtractive Hybridization (For RPF Samples only)

1. Resuspend the subtractive hybridization oligos in 1× TE to make 100 µM stocks. Then, prepare a 10 µM pool of oligos by adding 4 µl of each oligo to 24 µl of nuclease-free water.

2. In a 0.2 mL thin-walled tube, combine 5 µl of circularized cDNA with the following reagents (below).

Reagent	Per reaction
10 µM Oligo pool	2 µl
20× SSC	2 µl
Nuclease-free water	21 µl

3. In a thermocycler, heat the sample for 90 s at 95 °C and cool to 37 °C by cooling 3 °C per minute.
4. Prepare Subtractive hybridization buffer (*see* Table 8)
5. To prepare 1× bind/wash buffer, dilute the subtractive hybridization buffer 1:2 with nuclease-free water.
6. Meanwhile, for each sample, pipette 40 µl of MyOne Streptavidin C1 beads into a 1.5 mL tube.
7. Place the tube on a DynaMag™—2 Magnet and incubate for 1 min. Remove and discard the supernatant.
8. Add one volume of Bind/Wash buffer to the beads. Remove the tube from the magnet and vortex briefly. Put the tube back on the magnet and incubate for 1 min. Remove and discard the supernatant.
9. Repeat **step 8** twice.
10. Resuspend the beads in 80 µl Subtractive hybridization buffer.
11. Add 8 µl of 10 mg/mL yeast tRNA to each tube of washed beads, then add 72 µl of nuclease-free water to each tube. This step blocks the beads to reduce nonspecific binding of nucleic acids.
12. Place the tubes of beads on a tube rotator, and incubate for 20 min at room temperature.
13. Place the tubes on the DynaMag™—2 Magnet and incubate for 1 min. Remove and discard the supernatant.
14. Add 100 µl of 1× Bind/Wash buffer to each tube. Remove the tube from the magnet and vortex briefly. Put the tube back on the magnet and incubate for 1 min. Remove and discard the supernatant.
15. Repeat **step 14** twice.
16. Resuspend the beads in 80 µl of Subtractive hybridization buffer.
17. Add 50 µl of nuclease-free water to each sample of cDNA.
18. Add 80 µl of the sample to the 80 µl of MyOne Streptavidin C1 beads.

19. Place the tubes containing the beads and samples on a tube rotator. Incubate at room temperature for 15 min.
20. Place the 1.5 mL tubes containing the beads on the DynaMag™—2 Magnet and incubate at room temperature for 1 min.
21. Pipette off approximately 160 µl of the supernatant into a new 1.5 mL tube.
22. For each sample, add 460 µl of nuclease-free water, 64 µl of 3 M NaCl, and 1.5 µl of GlycoBlue.™
23. Add 700 µl isopropanol to each tube and incubate at −80 °C for 30 min. Alternatively, incubate at −20 °C for 1 h or overnight.
24. To pellet the cDNA, centrifuge for 30 min at 4 °C at 20,000 × *g*.
25. Remove and discard supernatant (*see* **Note 12**).
26. Wash the pellet by adding 600 µl of 70 % ethanol. Centrifuge the samples for 15 min at 4 °C at 20,000 × *g*.
27. Resuspend pellets in 10 µl nuclease-free water. The sample can be stored at −20 °C. Alternatively, proceed to the next step.

3.13 PCR Amplification

1. In a 0.2 mL thin-walled tube, combine 1 µl of circularized cDNA with the following reagents:

Reagent	Per reaction
5× Phusion HF buffer	8 µl
Nuclease-free water	25.8 µl
10 mM dNTP mix	0.8 µl
10 µM Universal Forward PCR Primer	2 µl
Phusion® DNA polymerase	0.4 µl

2. Add 2 µl of barcoded reverse PCR primer to each sample.
3. In a thermocycler, use following PCR protocol 98 °C: (30 s) [98 °C (10 s), 64 °C (10 s), 72 °C (30 s)] × 12 to amplify the DNA (*see* **Note 13**).

3.14 PCR Product Purification and Analysis

1. Clean up the PCR product using a Zymo DNA Clean and Concentrator—5™ column, according to the manufacturers instructions. Elute in 10 µl TE.
2. Determine the size and concentration of the PCR product by running 1 µl on a D1000 ScreenTape on a Tapesation (or Bioanalyzer equivalent). The PCR product is around 160 bp, and can also be visualized on a native 8 % polyacrylamide gel if a tapestation or bioanalyzer is unavailable (*see* **Note 14**).

3.15 Data Analysis

PCR products must be subjected to Illumina high-throughput sequencing. At least 10–20 M reads should be collected for RNA-seq samples from microbial organisms, and 20–30 M for metazoans. Because they are often highly contaminated with rRNA, twice as many reads are typically required from RPF libraries. Raw RNA-seq and RPF sequence reads must be processed computationally to derive measures of mRNA translation efficiencies, ribosome loads, pileups and stalling. Often these measures are computed via ad hoc analyses. However, a unified framework of computing them can begin with computation of a ribosome profile: the sequence of counts *cti* for each codon *i* of transcript *t* where *cti* gives the number of measured footprints with the ribosome P-site at location *i*.

The major challenges for deriving this profile are dealing with short sequence reads that map to multiple loci, sequencing bias correction, and the choice of the P-site within the footprint, in addition to read processing for quality control. None of these steps have been standardized. Here, we describe Ribomap [25], an automatic pipeline that addresses the challenges listed above and outputs isoform-level ribosome profiles, ribosome loads, and translation efficiencies. Ribomap takes raw ribo-seq and RNA-seq reads as input, trims linker sequences, filters out rRNA contaminate reads and reads with the wrong size, aligns the reads to the transcriptome, estimates the transcript abundances from the RNA-seq reads, and produces the *cti* counts for each transcript and codon position where the positions of multi-mapped ribo-seq reads are guided by transcript abundance. It allocates the P-site for each ribo-seq read with read-length-specific offset and corrects for sequencing bias by normalizing a transcript's ribosome profile with its mRNA profile (as done in refs. 24 and 26). In addition, Ribomap provides sub-codon resolution, nucleotide-level ribosome footprint coverage profiles including the UTR regions. Furthermore, Ribomap also computes ribosome loads, translation efficiency, and the relative abundance for each transcript. Lastly, Ribomap reports transcripts in order of the rank difference between the relative transcript abundance and ribosome load to help identify isoforms with different translation efficiency.

The command to run Ribomap in Linux or Mac OS X is:

run_ribomap.sh --rnaseq_fq *rnaseq.fq.gz* --riboseq_fq *riboseq.fq.gz* --contaminant_fa *contaminant.fa* --transcript_fa *transcript.fa* --cds_range *cds_range.txt --offset offset.txt*

Ribomap leverages existing read-processing tools for several of its steps, and includes optimized parameters for each tool. Users who wish to vary these parameters should refer to the README.txt file included with Ribomap. Below, we outline each step of Ribomap.

Ribomap input includes FASTQ files of raw RNA-seq (command line argument --rnaseq_fq) and ribo-seq (command line argument --riboseq_fq) reads. These are preprocessed to remove

the 3′ linker sequence, trim poor quality bases, and remove reads smaller than a plausible footprint size. Only ribo-seq reads with size between 27 and 33 are kept.

Reads that can be mapped to rRNA, tRNA, snoRNA are frequent contaminants in ribosome profiling libraries. By first mapping to sequences of these molecules (command line argument --contaminant_fa), users can retain unmapped reads for downstream analysis. Ribomap then aligns the remaining reads to the transcriptome (command line argument --transcript_fa) to find all potential mapping locations. Ribomap does the read preprocessing and the read mapping via the STAR aligner [27].

The final step of Ribomap resolves multi-mapping of ribo-seq reads. This step takes in a CDS range file that gives the coding region for each transcript (command line argument --cds_range). CDS range files for several model organisms are available with Ribomap. This step also uses the ribo-seq and RNA-seq alignment bam files (produced automatically above), and the transcript abundance estimation file (produced automatically with the Sailfish [28] system for isoform abundance estimation). Alternatively, abundance estimations from eXpress [29] or Cufflinks [30] can be used if preferred and available. The Psite for a read is adjusted based on the read start from the ribo-seq alignment bam with a read-length-specific offset (command line argument --offset) provided by an offset text file specifying the included read length and its P-site offset. Using the multi-map resolved ribo-seq read loci, Ribomap outputs isoform-specific codon-resolution ribosome profiles, a nucleotide-level profile, and isoform-specific ribosome loads, relative mRNA abundance levels and translation efficiencies.

4 Notes

1. In our experience, translation is much more sensitive than mRNA abundance to variations in culture media and timing of sample preparation. Thus, replicates are essential for ribosome profiling experiments. We find YEPD medium must be prepared through separate autoclaving of YEP and Dextrose solutions to avoid uncontrolled differences in caramelization. Identical batches of media must be used in replicate experiments. Each 650 mL culture gives sufficient material for preparation of up to three ribosome footprint libraries.
2. We often use styrofoam shipping containers from life science suppliers.
3. This step uses vacuum filters to rapidly purify yeast from culture. While many laboratories have vacuum on the bench, these vacuums are sometimes weak. We use a dedicated vac-

uum pump to maintain a constant vacuum necessary for rapid filtration.

4. Because the timing between cycloheximide treatment and sample preparation is critical, do not add cycloheximide to multiple cultures simultaneously.
5. We weigh the centrifuge tubes with an analytical balance and carefully add sucrose cushion solution if needed.
6. Up to six sucrose cushions can be run simultaneously using the TLA-100.3 rotor.
7. If a thermomixer is unavailable, heat the samples to 65 °C in a heat block for 5 min, vortexing briefly once a minute.
8. These mRNA enrichment procedures invariably bias the population of transcripts being analyzed [31]. Notably, polyadenylated (polyA+) selection, in addition to failing to capture non-polyA+ transcripts, poorly captures transcripts with short or heterogeneous tails and may generate bias in favor of low-turnover transcripts with homogeneous tail lengths [32, 33]. Though generally considered superior to polyA selection, rRNA depletion methods do introduce some bias [34] and may complicate comparisons of RNA-seq and RPF libraries by skewing the ratio of protein coding to noncoding reads [33]. Therefore, the choice of mRNA enrichment method depends on the nature of the study and its sensitivity to bias in the transcript pool.
9. In the absence of cycloheximide, a substantial number of ribosomes protect 22 nucleotide long mRNA fragments [20]. As such, researchers working without cycloheximide should consider purifying 20–32 nucleotide fragments instead.
10. We often place the closed conical tube in a beaker containing warm water to hasten urea dissolution.
11. Polymerized gels should be free of air bubbles and should appear smooth (without wrinkles or obstructions). Gels can be stored at room temperature overnight by wrapping any exposed surfaces in paper towels soaked with 1× TBE, followed by saran wrap to prevent drying. Do not cool the gel, as the urea will precipitate.
12. At this point care must be taken to not pipette off the pellet or dislodge the pellet.
13. For most samples, 1 μl of template is sufficient to generate enough PCR product after 12 cycles of amplification. However, for lower concentration samples, increasing the amount of template to 2–6 μl, and increasing the total reaction volume to 50–60 μl may be required. For high concentration samples, too much template will result in secondary PCR products and a reduction in the amount of desired PCR product. This phenomenon is likely a result of the polymerase continuing along

a circular template, resulting in a "smear" of products larger than ~160 bp peak of products. If this occurs, reduce the amount of starting template by diluting the original template. A dilution rage in the range of 1:2 to 1:8 should result in the desired PCR product.

14. Empty library, which results from amplification of residual RT oligo is an undesirable product that is ~140 base-pairs in length. The empty library may occur if the cDNA was not properly resolved during PAGE purification. To avoid high-throughput sequencing of empty library, it is advisable to remove this product if it represents more than 10 % of the library. It is possible size select and purify the PCR products on an agarose gel. Then, reanalyze the purified product for size and concentration.
15. The 6 nucleotide barcode sequences should match standard Illumina barcodes, which can be found online or in consultation with a high-throughput sequence service provider. Illumina sequencers require "balanced" sequences to deconvolute barcoded libraries. In general, the pool of barcodes needs to have at least one A or C and at least one G or T at each position, e.g., GCCAAT and CTTGTA.

Acknowledgements

This work has been partially funded by the US National Science Foundation (CCF-1256087, CCF-1319998) and US National Institutes of Health (R21HG006913 and R01HG007104).

References

1. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
2. Ingolia NT (2010) Chapter 6 - Genome-wide translational profiling by ribosome footprinting. In: Guide to yeast genetics: functional genomics, proteomics, and other systems analysis, vol 470, 2nd edn. Academic, New York, pp 119–142
3. Brar GA, Rouskin S, McGeachy AM et al (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 7:1534–1550
4. Ingolia NT, Brar GA, Rouskin S et al (2013) Genome-wide annotation and quantitation of translation by ribosome profiling. Curr Protoc Mol Biol. doi:10.1002/0471142727.mb0418s103
5. Gerashchenko MV, Lobanov AV, Gladyshev VN (2012) Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proc Natl Acad Sci U S A 109:17394–17399
6. Vaidyanathan PP, Zinshteyn B, Thompson MK, Gilbert WV (2014) Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA 20:912–922
7. Zid BM, O'Shea EK (2015) Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature 514:117–121
8. Bazzini AA, Lee MT, Giraldez AJ (2012) Ribosome profiling shows that miR-430

reduces translation before causing mRNA decay in zebrafish. Science 336:233–237

9. Stadler M, Fire A (2013) Conserved translatome remodeling in nematode species executing a shared developmental transition. PLoS Genet 9:e1003739
10. McManus J, May GE, Spealman P, Shteyman A (2013) Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. doi:10.1101/gr.164996.113
11. Artieri CG, Fraser HB (2013) Evolution at two levels of gene expression in yeast. Genome Res. doi:10.1101/gr.165522.113
12. Li G-W, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484:538–541
13. Dunn JG, Foo CK, Belletier NG et al (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. eLife 2:e01179
14. Smith JE, Alvarez-Dominguez JR, Kline N et al (2014) Translation of small open reading frames within unannotated RNA transcriptsin *Saccharomyces cerevisiae*. Cell Rep 7:1858–1866
15. Schneider-Poetsch T, Ju J, Eyler DE et al (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Methods 6:209–217
16. Wolin SL, Walter P (1988) Ribosome pausing and stacking during translation of a eukaryotic mRNA. EMBO J 7:3559–3569
17. Gerashchenko MV, Gladyshev VN (2014) Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Res 42:e134
18. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802
19. Lee S, Liu B, Lee S et al (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 109: E2424–E2432
20. Lareau LF, Hite DH, Hogan GJ, Brown PO (2014) Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. eLife 3:e01257
21. Tuller T, Carmi A, Vestsigian K et al (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354
22. Larsson O, Sonenberg N, Nadon R (2011) ANOTA: analysis of differential translation in genome-wide studies. Bioinformatics 27: 1440–1441
23. Olshen AB, Hsieh AC, Stumpf CR et al (2013) Assessing gene-level translational control from ribosome profiling. Bioinformatics 29:2995–3002
24. Artieri CG, Fraser HB (2014) Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res 24:2011–2021
25. Wang H, McManus CJ, Kingsord C (2015) Isoform-level ribosome occupancy estimation guided by transcript abundance with Ribomap. http://www.cs.cmu.edu/~ckingsf/software/ribomap/
26. Zupanic A, Meplan C, Grellscheid SN et al (2014) Detecting translational regulation by change point analysis of ribosome profiling data sets. RNA 20:1507–1518
27. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21
28. Patro R, Mount SM, Kingsford C (2014) sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol. doi:10.1038/nbt.2862
29. Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 10: 71–73
30. Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578
31. Raz T, Kapranov P, Lipson D et al (2011) Protocol dependence of sequencing-based gene expression measurements. PLoS One 6:e19287
32. Chang H, Lim J, Ha M, Kim VN (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell 53:1044–1052
33. Sultan M, Amstislavskiy V, Risch T et al (2014) Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics 15:675
34. Adiconis X, Borges-Rivera D, Satija R et al (2013) Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods 10:623–629

Chapter 6

Studying Isoform-Specific mRNA Recruitment to Polyribosomes with Frac-seq

Rocio T. Martinez-Nunez and Jeremy R. Sanford

Abstract

Gene expression profiling is widely used as a measure of the protein output of cells. However, it is becoming more evident that there are multiple layers of post-transcriptional gene regulation that greatly impact protein output (Battle et al., Science 347:664–667, 2014; Khan et al., Science 342:1100–1104, 2013; Vogel et al., Mol Syst Biol 6:400, 2010). Alternative splicing (AS) impacts the expression of protein coding genes in several ways. Firstly, AS increases exponentially the coding-capacity of genes generating multiple transcripts from the same genomic sequence. Secondly, alternatively spliced mRNAs are subjected differentially to RNA-degradation via pathways such as nonsense mediated decay (AS-NMD) or microRNAs (Shyu et al., EMBO J 27:471–481, 2008). And thirdly, cytoplasmic export from the nucleus and translation are regulated in an isoform-specific manner, adding an extra layer of regulation that impacts the protein output of the cell (Martin and Ephrussi, Cell 136:719–730, 2009; Sterne-Weiler et al., Genome Res 23:1615–1623, 2013). These data highlight the need of a method that allows analyzing both the nuclear events (AS) and the cytoplasmic fate (polyribosome-binding) of individual mRNA isoforms.

In order to determine how alternative splicing determines the polyribosome association of mRNA isoforms we developed Frac-seq. Frac-seq combines subcellular fractionation and high throughput RNA sequencing (RNA-seq). Frac-seq gives a window onto the translational fate of specific alternatively spliced isoforms on a genome-wide scale. There is evidence of preferential translation of specific mRNA isoforms (Coldwell and Morley, Mol Cell Biol 26:8448–8460, 2006; Sanford et al., Genes Dev 18:755–768; Zhong et al., Mol Cell 35:1–10, 2009; Michlewski et al., Mol Cell 30:179–189, 2008); the advantage of Frac-seq is that it allows analyzing the binding of alternatively spliced isoforms to polyribosomes and comparing their relative abundance to the cytosolic fraction. Polyribosomes are resolved by sucrose gradient centrifugation of cytoplasmic extracts, subsequent reading and extraction. The total mRNA fraction is taken prior ultracentrifugation as a measure of all mRNAs present in the sample. Both populations of RNAs are then isolated using phenol–chloroform precipitation; polyadenylated RNAs are selected and converted into libraries and sequenced. Bioinformatics analysis is then performed to measure alternatively spliced isoforms; several tools can be used such as MISO, RSEM, or Cufflinks (Katz et al., Nat Methods 7:1009–1015, 2010; Li and Dewey, BMC Bioinformatics 12:323, 2011; Trapnell et al., Nat Protoc 7:562–578, 2012). Comparison of total mRNAs and polyribosome-bound mRNAs can be used as a measure of the polyribosome association of specific isoforms based on the presence/absence of specific alternative splicing events in each fraction. Frac-seq shows that not all isoforms from a gene are equally loaded into polyribosomes, that mRNA preferential loading does not always correlate to its expression in the cytoplasm and that the presence of specific events such as microRNA binding sites or Premature Termination Codons determine the loading of specific isoforms into polyribosomes.

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_6,

Key words RNA-seq, Polyribosome, Subcellular fractionation, Posttranscriptional regulation, Alternative splicing, Translation

1 Introduction

Polyribosome fractionation is a well-established and powerful technique that has been employed in multiple organisms such as plants, yeast, fly, and mammalian cells [1–4]. Not only does it provide information of the mRNAs actively engaged in translation, but it also gives information about the global translational status of a cell: as an example, cells in translational arrest show an increase in the 80S fraction due to translation initiation inhibition. It has been suggested that polyribosome profiling of mammalian cells presents more challenges due to their more unstable nature [5]. Importantly, mRNA association with polyribosome provides a more accurate reflection of the protein levels than steady state mRNA analysis [6] and thus analyzing only mRNA levels does not reflect the translational status of mRNAs.

Post-transcriptional gene regulation is an essential mechanism in metazoans. Alternative splicing increases the coding potential of our genome and alters the translational fate of specific messages due to isoform-specific changes in mRNA export, degradation, and translational control. There is a poor correlation between protein levels and mRNA levels in eukaryotic systems [7, 8] and we hypothesized that alternative splicing may influence the translationability of mRNA isoforms. To test this hypothesis we developed Frac-seq, which integrates and compares mRNA isoform levels in different fractions of the cell: total and polyribosome-bound mRNAs. Frac-seq allows analyzing alternative splicing effects on isoform-specific translation on a genome-wide scale.

Here we describe a method that successfully isolates individual polyribosomal populations from human cells employing sucrose gradients and ultracentrifugation (outline in Fig. 1). Cells are lysed and a fraction of the lysate is separated for total mRNA isolation (as a readout of all isoforms present in the cell). Lysates are loaded onto sucrose gradients and ultracentrifuged, allowing for separation of distinct polyribosome populations. RNA from the total fraction and the polyribosomal fractions is then isolated using phenol–chloroform extraction, converted into libraries and sequenced. Bioinformatics analysis allows characterization and comparison of all isoforms present in the cells and those preferentially loaded into polyribosomes employing pipelines such as MISO, RSEM, or Cufflinks [9–11]. This RNA-seq data can be employed to pinpoint the exact isoforms that have the potential (total mRNA) and those that contribute (polyribosome-bound mRNA) to the proteome, as well as downstream analysis of RNA elements that contribute to preferential loading into polyribosomes.

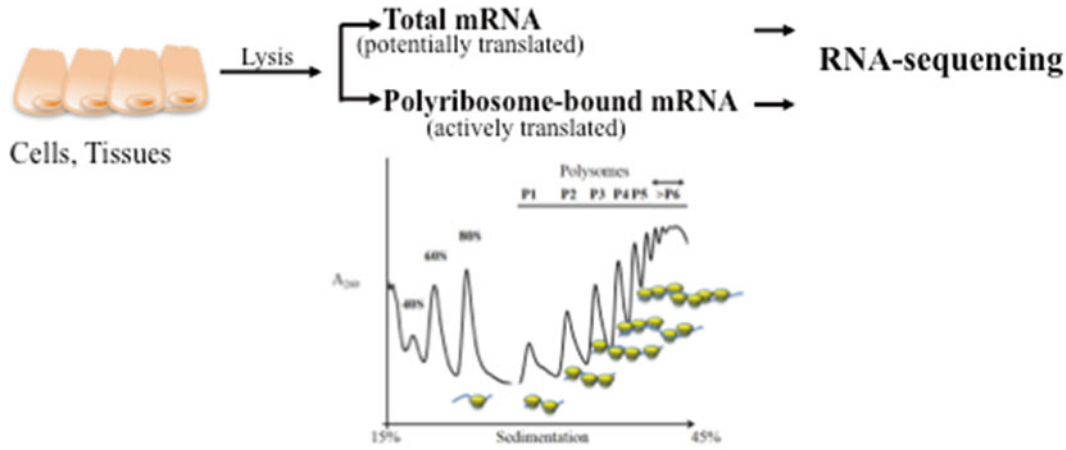

Fig. 1 Schematic of Frac-seq

2 Materials

Always pass stock solutions through a 0.22 μm filter.

1. Polyribosome Gradient Buffer (PGB): 20 mM Tris–HCl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$ (*see* **Note 1**). Stock solutions (highest and lowest % sucrose) are prepared by dissolving sucrose in PGB (*see* **Note 2**). Sucrose solutions can be aliquoted and stored at 4 °C for several months and at –20 °C for longer.
2. Lysis buffer (LB): 0.5 % NP-40, 20 mM Tris–HCl pH 7.5, 100 mM KCl, 10 mM $MgCl_2$ and protease inhibitors (*see* **Note 3**).
3. Chloroform, co-precipitant such as glycogen (*see* **Note 4**), 100 % ethanol, 75 % ethanol to perform RNA extraction using phenol–chloroform extraction from both total lysate and the individual polyribosome fractions.
4. SDS 10 % Buffer: 10 % SDS, 500 mM EDTA, 1 M Tris–HCl pH 7.5.
5. Open top tubes for ultracentrifugation (Seton Tubes 151-514B for SW40 rotor and 151-514A for SW41 rotor).
6. Ultracentrifuge with swinging buckets (Beckman SW40, Beckman SW41).
7. Cycloheximide.
8. Phosphate buffer saline (PBS) without Ca^{2+} or Mg^{2+}.
9. Gradient Station (Biocomp Instruments, New Brunswick Canada).

3 Methods

It is very important to underline that all buffers and solutions must be filtered prior to their use. This avoids the presence of small particles that may disrupt the gradients and/or affect the fractionation of the sample.

We employ a Gradient Station (Biocomp) in order to prepare the gradients, measure the absorbance at 260 nm across the gradient and extract the polyribosome populations. This system presents several advantages: ease of use, homogenous gradient formation, flexibility with different types and sizes of rotors/buckets, minimal disruption of the sample when determining the A_{260}, and precise isolation of polyribosome populations avoiding contamination with sample between fractions [12].

Additionally we recommend that any solutions containing EDTA are kept away when performing polyribosomal fractionation (*see* **Note 5**).

3.1 Sucrose Gradients Preparation

Bring the sucrose stock solutions to room temperature.

Mix the stock solutions following the manufacturer's instructions and place them in the Gradient Station, select appropriate program according to the % sucrose and run. Formed gradients must be chilled at 4 °C until their use. Avoid any shaking of the formed gradients as this will greatly impact on polyribosome separation.

When possible, prepare duplicated samples. Depending on the system, one sample may need to be "sacrificed" for scanning the A_{260} and mapping of the fractions. Some systems allow for reading and extraction within the same sample (*see* **Note 6**).

3.2 Cell Lysis

It is very important to carry all steps on ice to minimize RNA and protein degradation and to maintain integrity of the polyribosomes during isolation.

1. Place LB and tubes where extracts will be lysed (1.5 mL tubes) on ice.
2. Incubate cells in the presence of 100 μg/mL cycloheximide during 10 min at 37 °C 5 % CO_2.
3. Wash cells three times with ice-cold PBS. Carefully remove all trace of PBS from the sample.
4. Place cells on ice. Add prechilled LB to washed cells. Cell numbers may vary; we have successfully fractioned from 3 million cells up to 15 million cells. Adjust the amount of buffer depending on the number of cells, maximizing the protein concentration of the lysates. Typically, add 500 μL of LB to a 15 cm^2 dish.
5. Pass lysates three times through a 23G needle avoiding foaming. Incubate lysates on ice for 7–10 min (depending on the amount of material used as well as the cell type).

6. Centrifuge extracts at 7,200 to 11,200 × *g* in a bench centrifuge at 4 °C. The pellet contains most nuclear fraction as well as membranes. Avoid contamination from the pellet in the subsequent steps. It is very important that the lysate is clear from all precipitants, as the presence of any particle will disrupt the gradient and fractionation.

3.3 Gradient Loading and Ultracentrifugation

1. Prechill the SW41 buckets as well as the formed sucrose gradients where the samples are going to be ultracentrifuged. Place buckets with sucrose gradient on ice prior to loading (*see* **Note 7**).
2. Aliquot 10 % of the lysate for total mRNA isolation (keep in TRIzol LS or SDS 10 % Buffer in −80 °C until use).
3. Carefully load the lysate on top of the gradient without disrupting the surface of the gradient. The range for long caps in SW40 or SW41 rotors goes from 200 to 400 μL. We recommend loading as much as possible, leaving 1–2 mm at the top to avoid spillage.
4. We employed 40,000 rpm (285,000 × *g* at r maximum) during 1 h and 20 min for SW41 rotors and 1 h and 40 min for SW40 rotors. The best results are achieved combining fastest speed centrifugation and shortest times (*see* **Note 8**). Centrifuge always at 4 °C with maximum acceleration and maximum break, prechilling the ultracentrifuge to obtain best results.

3.4 Scanning Polyribosomes and Fractionation

We monitor the A_{260} of fractionated ribosomes using the Gradient Station, either using the BioRad UV Monitor (EM-1) or the Bioprobes (Biocomp). When using the EM-1 system, two gradients per sample must be prepared: one for scanning and mapping ribosomal peaks and another to extract the individual polyribosomal populations. The use of the BioProbes allows scanning and isolation from the same gradient, if sample material is a limiting step. They are also more sensitive than the EM-1 system.

After ultracentrifugation, the buckets containing the sucrose gradients must be placed on ice. Scan the sucrose gradient as soon as possible, within 10 min after the centrifugation has finished to avoid diffusion of material between polyribosome populations. An example of a scan of proliferating cells and non-proliferating cells is shown in Fig. 2a, b, respectively.

Fractionation can be continuous or noncontinuous. Continuous fractionation consists in extraction of serial aliquots of the gradient, from top to bottom (schematic in Fig. 2c).

Noncontinuous fractionation consists in isolation of individual polyribosomal peaks with no mixing between them or with non-polyribosomal populations (schematic in Fig. 2d).

For pros and cons, *see* **Note 9**.

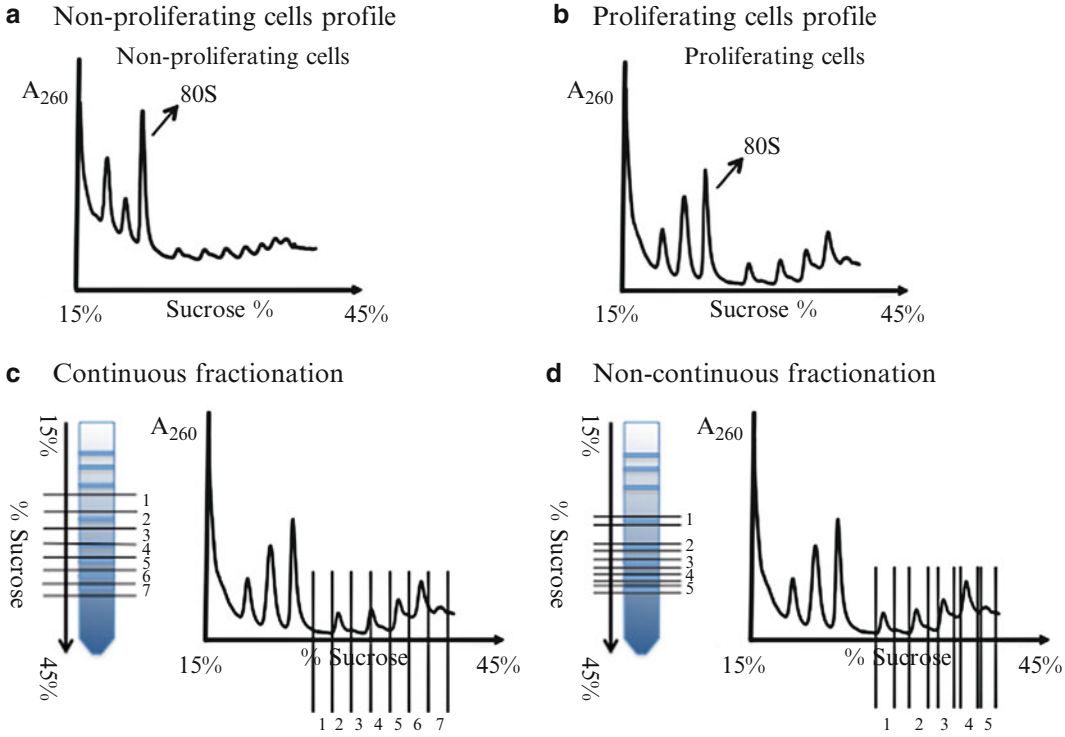

Fig. 2 Polyribosome profiles. (**a**) Polyribosome profile of non-proliferating cells. Note the increased 80S peak and smaller polyribosome populations. (**b**) Polyribosome profile of proliferating cells. Note the smaller 80S and bigger polyribosome populations. (**c**) Schematic of continuous fractionation. Fractions are collected continuously from top to bottom of the gradient, mixing polyribosome populations and material between polyribosome peaks. (**d**) Schematic of non-continuous fractionation. Polyribosome populations are individually extracted avoiding mixing with other fractions

3.5 RNA Isolation, Reverse Transcription (RT) and PCR

We have employed both TRIzol LS and acid phenol–chloroform extraction methods with comparable results. The bands containing polyribosomal fractions tend to be ~250 μL of material and specifications below are adjusted volumes for that starting material. There are many protocols available for these methods and we have outlined here the ones we have successfully used.

We recommend extraction of individual fractions from each gradient, quantitation of RNA from each fraction (Nanodrop) and determination of RNA integrity using a Bioanalyzer before performing reverse Transcription of individual fractions. The cDNA of each individual fraction can be then assayed to map the localization of specific isoforms in different polyribosomal populations [12]. If assaying comparisons between total and polyribosome-bound mRNAs, mix the different cDNA obtained from individual fractions and perform PCR on the mixture, for more robust results.

3.5.1 Protocol for TRIzol LS

For RNA extraction using TRIzol LS, add 750 μL of TRIzol LS, vortex for 15 s, and leave for 5 min at room temperature. Add 200 μL chloroform, vortex for 20 s, leave for 5 min at room temperature, and spin at 11,200 × *g* using a benchtop centrifuge for 15 min at 4 °C to separate phases. Carefully take aqueous phase containing RNA and place it in a clean tube, adding half its volume of isopropanol. Add 15–30 μg of glycogen to it, vortex for 20–30 s, and leave for 10 min at room temperature. Place samples at −80 °C for a minimum of 1 h (longer times allow precipitating more material, use over night incubation if necessary). Centrifuge at 11,200 × *g* using a benchtop centrifuge for 20–30 min at 4 °C to pellet RNA with co-precipitant. Carefully remove supernatant and wash pellet with 75 % ice-cold ethanol. Do not disrupt the RNA pellet. Leave on ice for 10–15 min and centrifuge again at 11,200 × *g* using a benchtop centrifuge for 10 min at 4 °C. Discard ethanol and repeat the washing step. Remove all excess of ethanol and air-dry the pellet. Resuspend in RNAse/DNAse-free water (20 μL per fraction and 25 μL for total RNA).

3.5.2 Protocol for Acid Phenol–Chloroform

Add 28 μL of SDS 10 % Buffer. Vortex for 15 s. Add 28 μL of 0.3 M NaAc pH 5.2 and 500 μL Vol phenol–chloroform–IAA. Vortex for 1 min and centrifuge at max speed for 2 min at room temperature. Remove aqueous phase and add to it 2× Vol chloroform. Vortex for 15–20 s and centrifuge at max speed for 2 min at room temperature. Remove aqueous phase, add 2.5 Vol. 100 % ethanol and 15–30 μg of glycogen, vortex for 20–30 s, and leave for 10 min at room temperature. Place at −80 °C for a minimum of 1 h (longer times allow precipitating more material, use over night incubation if necessary). Pellet RNA at max speed for 10 min. Wash twice with 75 % ice-cold ethanol. Air-dry the RNA pellet and resuspend in RNAse/DNAse-free water (20 μL per fraction and 25 μL for total RNA).

3.6 RNA-Sequencing and Bioinformatics

There are many platforms available for RNA-seq. We do not intend here to compare these different platforms; however, there are certain parameters that we advise to take into account when performing Frac-seq:

- Take advantage of biological replicates within your experimental design. We define replicates as cytosolic extracts prepared from different plates of cells or tissues. Replicates are essential for estimating a false discovery rate.
- Use paired-end sequencing to allow detection of alternatively spliced isoforms;
- Longer reads (50–100 bp) tend to give more accurate information about exon-junctions and thus the mapping and accuracy of isoform detection is better than shorter reads.

3.6.1 Mapping

There are many different tools available for the mapping [13, 14]. Ideally the best decision regarding the mapping is made after performing a simulation of the data and comparing different mapping tools.

3.6.2 Alternative Splicing Analysis

Similarly, there are several options and programs available to perform alternative splicing analysis [9–11]. We employed MISO [9], a Bayesian inference method that computes the probability of a read originating from a specific isoform. We considered Psi (ψ Percentage Spliced In) of eight possible alternative splicing: alternative 5′splice sites(A5SS), alternative 3′ splice sites (A3SS), skipped exons (SE), mutually exclusive exons (MXE), alternative first exons (AFE), alternative last exons (ALE), retained introns (RI) and tandem 3′UTRs (alternative polyadenylation). We used the Bayes Factor (BF) as a measure of biological reproducibility of the different inclusion of splicing events between cytoplasm and polyribosomes. We calculated $\Delta\psi$ ($\psi_{Polysome} - \psi_{Total}$) for each individual isoform as a measure of difference presence between polyribosomal binding and total cytoplasmic presence. By plotting the Spearman's rank correlation coefficient between biological replicates of $\Delta\psi$ and the Bayes Factor (BF, taking $BF > 1$), we observed a very good correlation at BF of five [12]. These data show that the BF may be used as a measure of biological reproducibility. We recommend that this cutoff BF is recalculated when performing Fracseq as different biological systems, RNA-seq pipelines and data processing may have differences in reproducibility. Results can be then subdivided into alternative splicing events ($0.1 < \psi < 0.9$) that show difference or no difference between cytoplasmic and polyribosome presence ($|\Delta\psi| < 0.1$, $BF < 1$). Validation of the events by RT-PCR in the cytoplasmic fraction and polyribosomal-bound fractions should be used to calculate FDR (False Discovery Rates).

4 Notes

1. The concentration of $MgCl_2$ is key to the stability of the polyribosomal populations. We recommend a pre-run employing two or three concentrations of $MgCl_2$ ranging from 10 to 100 mM in order to optimize the fractionation.
2. We have successfully fractionated human cell extracts employing two types of sucrose gradients: 10–50 % and 15–45 %. We recommend a preliminary run with these two types of gradients to determine the optimal % sucrose that best separates the polyribosome populations.
3. Depending on the cell type, LB may need the addition of 500 μg/mL cycloheximide and/or RNAse inhibitor. Certain cell types and/or tissues are rich in RNAse enzymes that can rapidly degrade the sample RNA, rendering it useless. Always

add protease inhibitors such as Pefabloc to LB. Add these fresh every time just prior to cell lysis.

4. We recommend checking the protocol of library preparation as some co-precipitants used during RNA extraction are incompatible with certain methods. Glycogen is widely accepted in most library preparation protocols and yields suitable amounts of RNA.
5. Most of trypsin solutions to detach adherent cells contain EDTA. Preferably use EDTA-free solutions when passing adherent cells that will be used for polyribosome fractionation.
6. The use of BioProbes (Biocomp) allows for scanning and fractionation of the same sample with no mixing between samples. They consist on two needles that scan the absorbance at 260 nm and/or 280 nm from top to bottom of the gradient. This minimal mixing is observable when the same sample is read twice, creating identical profiles, by only rotating the tube 90° between the two scans.
7. We have successfully fractionated lysates employing SW40 Ti and SW41 Ti swinging buckets (Beckman Coulter). Other types of swinging rotors/buckets may be suitable.
8. Centrifugation times may be adjusted depending on the readout of the gradients. If most fractions accumulate at the bottom of the gradient, employ longer times. On the other hand, if no separation is observed, employ shorter times.
9. Continuous fractionation allows recovering more material but it is not as precise as fractions will contain impurities between the polyribosomal populations.

Noncontinuous fractionation allows recovering individual polyribosomal populations. The material recovered may be less, but it is more pure.

Depending on the downstream applications one may use one or another. For detection of isoform specific binding to polyribosomes, noncontinuous fractionation is the method to employ, as it allows precise mapping of specific isoforms in individual polyribosomal populations.

References

1. Penman S et al (1963) Polyribosomes in normal and poliovirus-infected Hela cells and their relationship to messenger-RNA. Proc Natl Acad Sci U S A 49:654–662
2. Atkin AL et al (1995) The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm. Mol Biol Cell 6:611–625
3. Satterfield TF, Pallanck LJ (2006) Ataxin-2 and its Drosophila homolog, ATX2, physically assemble with polyribosomes. Hum Mol Genet 15:2523–2532
4. Zanetti ME et al (2005) Immunopurification of polyribosomal complexes of Arabidopsis for global analysis of gene expression. Plant Physiol 138:624–635
5. Esposito AM et al (2010) Eukaryotic polyribosome profile analysis. J Vis Exp 2010:1948

6. Ingolia NT et al (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 7: 1534–1550
7. Kislinger T et al (2006) Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. Cell 125:173–186
8. Griffin TJ et al (2002) Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol Cell Proteomics 1:323–3339
9. Katz Y et al (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 7:1009–101510
10. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:32311
11. Trapnell C et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578
12. Sterne-Weiler T et al (2013) Frac-seq reveals isoform-specific recruitment to polyribosomes. Genome Res 23:1615–162313
13. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
14. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

Chapter 7

Use of the pBUTR Reporter System for Scalable Analysis of 3′ UTR-Mediated Gene Regulation

Arindam Chaudhury and Joel R. Neilson

Abstract

Posttranscriptional control of mRNA subcellular localization, stability, and translation is a central aspect of gene regulation and expression. Much of this control is mediated via recognition of a given mRNA transcript's 3′ untranslated region (UTR) by microRNAs and RNA-binding proteins. Here we describe how a novel, scalable *piggyBac*-based vector, *pBUTR*, can be utilized for analysis of 3′ UTR-mediated posttranscriptional gene regulation (PTGR) both in vitro and in vivo. This vector is specifically designed to express a selection marker, a control reporter, and an experimental reporter from three independent transcription units. Expression of spliced reporter transcripts from medium-copy non-viral promoter elements circumvents several potential confounding factors associated with saturation and stability, while stable integration of these reporter and selection elements in the context of a DNA transposon facilitates experimental reproducibility.

Key words Posttranscriptional gene regulation, PTGR, 3′-UTR, pBUTR, *piggyBac*, Reporter, miRNA sensors, RNA-binding proteins, mRNA stability

1 Introduction

Coordinated regulation of gene expression is fundamentally important for all aspects of cellular function. Historically, the most widely utilized practice in assessing coordinated regulation of gene expression has been via analysis of mRNA steady-state expression using either microarray [1] or next-generation sequencing approaches [2, 3]. Both approaches provide powerful information about genome wide changes in transcript abundance. However, these approaches fail to provide any information in regard to whether mRNA that has been transcribed is indeed being actively utilized by the translation machinery to produce protein. Emerging evidence strongly suggests that regulation of gene expression at the translational level contributes as much, if not more, to gene expression than transcription [4–6]. In fact, a reasonable amount of evidence suggests that coordinated changes in posttranscriptional

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_7,

regulatory networks occur during cellular differentiation and/or response to stimulus, and that these networks may profoundly alter cellular phenotype and behavior [7–10].

A significant amount of the control of mRNA subcellular localization, translation, and stability is mediated via *cis*-regulatory elements in the 3′ untranslated region (UTR) of target transcripts. These elements may be recognized by specific microRNAs and RNA-binding proteins [8, 11, 12]. Dysregulation of posttranscriptional control by microRNAs and RNA-binding proteins underlies distinct steps of pathogenesis in a wide spectrum of human diseases [12]. In some systems, 3′ UTR identity is itself sufficient to confer appropriate temporospatial gene expression in vivo [11]. Both alternative splicing [13] and alternative cleavage and polyadenylation [14] can alter 3′ UTR identity, and thus the visibility of related gene products to the posttranscriptional regulatory machinery. However, as compared to other facets of gene regulation, the contributions of these phenomena to gene regulation remain largely unexplored. Given that mutations within the 3′ UTRs of certain genes can significantly impact human health [15–17], it is of great interest to determine if and how genomic variations within the 3′ UTR, uncovered via genome-wide association studies and next-generation sequencing surveys, impact the pathology of the disease or phenotype with which they are associated. It is for this reason that we were motivated to develop a scalable and robust reporter system explicitly designed to model 3′ UTR-mediated regulation.

To these ends, we engineered a novel, scalable *piggyBac* transposon-based reporter system that we have named *pBUTR* (*piggyBac-based* 3′ UnTranslated Region reporter) [18]. We chose a DNA transposon-based system in consideration of our specific purpose. The integration of the *pBUTR* vector into the DNA of the target cell is essentially a stable transfection, which is generally superior to transient transfection in regards to experimental reproducibility and reduction of "noise." That the vector is DNA based allows for the inclusion of multiple independent transcription units. Thus, a control reporter may be expressed completely independently of the experimental reporter. Also, the use of a DNA-based vector allows the inclusion of splice junctions in each of the transcription units encoded within the vector. It is well established that transcripts that have not undergone splicing do not express as well as those that have, due in some part to the fact that the lack of an exon junction complex (EJC) marks unspliced transcripts as substrates for the nonsense-mediated decay (NMD) pathway [19]. Retro- and lentiviral vector systems do not have these features, and thus often times include a stability sequence such as the woodchuck hepatitis virus post-translationally regulated element (WPRE) element [20], which would be expected to confound native post-transcriptional regulation. In addition to these

limitations, the long terminal repeat (LTR) elements of retro- and lentiviral vectors may be recognized as foreign and silenced by the cell over time, a problem that is particularly observed in the context of transgenesis [21–23].

The *pBUTR* vector system is comprised of three independent transcription units—a G418 selection cassette, a control turboGFP reporter gene driven by *PGK* promoter, and a Gateway® [24] recombineering cassette under the control of the ubiquitin C (*UBC*) promoter (Fig. 1). These promoters were chosen because they drive expression at low-to-medium levels, and therefore are less likely to overwhelm any endogenous regulatory mechanisms. The *pBUTR* destination vector is generated via four-part Gateway® recombineering using an *att*L1/L2-flanked coding sequence of interest, an *att*R2/*att*L4 flanked 3′ UTR element, and an *att*R4/*att*L5-flanked minimal polyadenylation sequence [25] followed by a unique 24-nucleotide barcode. Upon recombination of these three elements into the parent vector, a bi-fluorescent reporter is produced that can be employed in both in vitro and in vivo model systems.

The *pBUTR* vector was functionalized with Gateway® technology to allow high-dimensionality screening and validation applications. Given that Gateway® recombineering is scalable—meaning

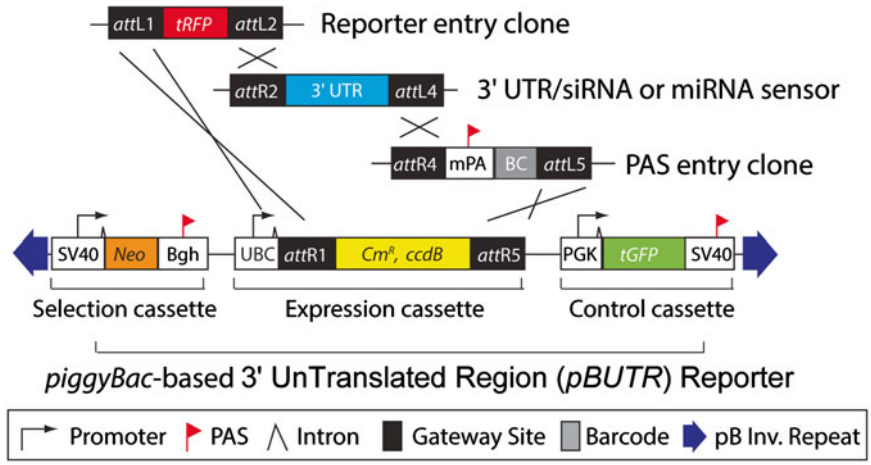

Fig. 1 Schematic representation of the *pBUTR* vector. The *pBUTR* destination vector is functionalized by four-part Gateway® recombineering using an *att*L1/L2-flanked tRFP (can be substituted with any coding sequence of interest), an *att*R2/*att*L4 flanked 3′ UTR element, and an *att*R4/*att*L5-flanked minimal polyadenylation sequence followed by a unique 24-nucleotide barcode. The inclusion of unique barcode elements with the minimal polyadenylation signal was made to allow analyses within pooled cell populations via flow cytometry and cell sorting. *att*XN, Gateway® recombination site; *tRFP*, turboRFP; *UTR*, untranslated region; *mPA*, minimum polyadenylation signal; *BC*, 24 nt barcode; *PAS*, polyadenylation signal; *SV40 (left)*, SV40 early promoter region. *Neo*, neomycin resistance gene; *Bgh*, bovine growth hormone polyadenylation signal; *UBC*, ubiquitin C promoter element; *CmR*, chloramphenicol-resistance gene; *PGK*, murine phosphoglycerate kinase 1 promoter; *tGFP*, turboGFP; *SV40*, SV40 late polyadenylation signal. Features not to scale

multiple individual 3′ UTR elements can be cloned into the vector in bulk—an inclusive, aggregate set of 3′ UTRs of interest can be rapidly generated and tested for contextual regulatory activity in pooled or arrayed format. Here we discuss how the *pBUTR* reporter can be used to study 3′ UTR-mediated gene regulation in vitro (in both arrayed and pooled format) and in vivo.

The E-cadherin transcriptional repressors *ZEB1* and *ZEB2* play established roles in epithelial to mesenchymal transition, both during tumor metastasis and during embryogenesis [26]. The mRNA transcripts of both of these gene products are characterized by multiple, validated miR-200 family recognition elements in their respective 3′ UTRs [26]. Cells with an epithelial phenotype express high relative levels of the miR-200b microRNA, which enforces posttranscriptional repression of the *ZEB1* and *ZEB2* mRNA transcripts. However, as cells undergo EMT, for example in response to transforming grown factor-beta (TGF-β), relative levels of miR-200b are reduced, allowing increased expression of ZEB1 and ZEB2 proteins and transcriptional repression of the *CDH1* (E-cadherin) gene. Previously described [26] wild-type and mutant (where each miR-200b-binding site has been ablated via site-directed mutagenesis) *ZEB2* 3′ UTR elements were recombineered into the *pBUTR* destination vector so as to confer regulation upon *tRFP* expression in the assembled reporter. We initially discuss how to study microRNA (miR-200b in this case)-mediated repression in a cell-based model of epithelial to mesenchymal transition (EMT), and then how to assess this regulation in vivo during embryogenesis. Entirely similar strategies can be employed to use the *pBUTR* reporter to study in vitro and in vivo 3′ UTR-mediated gene regulation in the context of siRNA/microRNA sensor activity, and posttranscriptional gene regulation (PTGR) by RNA-binding proteins, in both arrayed and pooled screening approaches.

2 Materials

2.1 BP Recombination Reaction

1. *att*B-flanked PCR products:
   (a) Turbo-RFP (tRFP—Evrogen) amplified with:
   *att*B1-tRFP-forward oligonucleotide primer –
   5′—GGGGACAAGTTTGTACAAAAAAGCAGGCTCGCCACCATGAGCGAGCTG—3′, and
   *att*B2-tRFP-reverse oligonucleotide primer –
   5′—GGGGACCACTTTGTACAAGAAAGCTGGGTAGATCCTACACATTGATCCTAGCAGAAGC—3′.
   (b) Amplify 3′ UTRs or siRNA/miRNA sensor elements with:
   *att*B2r-forward primer—5′—GGGGACCCAGCTTTCTT

GTACAAAGTGGT*NNNN…NNNN*—3′ and *att*B4-reverse primer—5′—GGGGACAACTTTGTATAGAAAA GTTGGGTG*NNNN…NNNN*—3′, where "N" is the specific priming sequence (*see* **Notes 1** and **2**). Note that the length of the specific priming sequence should be 18–21 nucleotides as for any other standard PCR amplification.

(c) Amplify the minimum polyadenylation/barcode element with *att*Br4-forward primer—5′-GGGGACAACTTTTCT ATACAAAGTTGAACTAGTAATAAAGG—3′ and *att*B5-reverse primer—5′—GGGGACAACTTTGTATA CAAAAGTTGCG- 3′ from a synthetic *att*Br4_mPA_ barcode_*att*L5 oligo: (ACAACTTTTCTATACAAAGTT GAACTAGTAATAAAGGATCCTTTATTTTCA T T G G A T C C G T G T G T T G G T T T T T T G T G T N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N N CGCAACTTTTGTATACAAA GTTGT), where the N sequence represents the 24-nucleotide barcode (*see* **Note 3**).

2. A *pDONR223 att*P1-attP2 plasmid to generate the *tRFP* entry clone, a *pDONR223 att*P2r-*att*P4 plasmid to generate the 3′ UTR or siRNA/miRNA sensor entry clone, and a *pDONR223 att*P4r-*att*P5 plasmid to generate the minimum polyadenylation/barcode entry clone.
3. BP Clonase II enzyme mix (*see* **Note 4**).
4. 2 μg/μl Proteinase K in 50 mM Tris, pH 8.0, 3 mM $CaCl_2$, 50 % Glycerol.
5. 1× TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, pH 8.0.
6. 37 °C water bath.
7. Vortex.
8. TOP10 competent *E. coli* cells.
9. LB agar plates containing spectinomycin (50 μg/ml).
10. Primers for colony PCR screening and sequencing: M13 forward (-20)—5′-GTAAAACGACGGCCAG-3′, M13 reverse—5′-CAGGAAACAGCTATGAC-3′.
11. One*Taq* DNA Polymerase or any other DNA polymerase with proof reading activity.
12. Agarose.
13. 10 μg/ml ethidium bromide in double-distilled water. Use at a final concentration of 0.5 μg/ml.
14. 1× Tris-acetate EDTA buffer: 40 mM Tris-acetate, 1 mM EDTA.

2.2 LR Recombination Reaction

1. Sequence verified *att*L1-*att*L2, *att*R2-*att*L4 and *att*R4-L5 donor plasmids.
2. *pBUTR* destination vector containing 5′ *att*R1 and 3′ *att*R5 sites.
3. LR Clonase II Plus enzyme mix (*see* **Note 4**).
4. 2 μg/μl Proteinase K in 50 mM Tris, pH 8.0, 3 mM $CaCl_2$, 50 % glycerol.
5. 1× TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA, pH 8.0.
6. 37 °C water bath.
7. Vortex.
8. One Shot Mach1 T1 chemically competent *E. coli* cells.
9. LB agar plates containing ampicillin (100 μg/ml) and kanamycin (100 μg/ml).
10. Primers for PCR screening: *UBC* forward—5′-ATTGTCC GCTAAATTCTGGC-3′, *PGK* reverse—5′- TAAAGCGCAT GCTCCAGAC -3′.
11. One*Taq* DNA Polymerase or any other DNA polymerase with proof reading activity.
12. Agarose.
13. 10 μg/ml ethidium bromide in double-distilled water. Use at a final concentration of 0.5 μg/ml.
14. 1× Tris-acetate EDTA buffer: 40 mM Tris-acetate, 1 mM EDTA.

2.3 Cell Culture, Transfection, and Stable Clone Generation for In Vitro Experiments

1. MCF10A cell line: Any appropriate adherent or suspension cell line can be similarly used.
2. Growth media for MCF10A cells: DMEM/F12 medium, 5 % horse serum, 0.01 mg/ml bovine insulin, 0.5 μg/ml hydrocortisone, 100 ng/ml cholera toxin, 20 ng/ml human EGF, and 100 U/ml penicillin and 0.1 mg/ml streptomycin.
3. Mature *pBUTR* vector.
4. Plasmid containing transposase (*pCMV-HA-m7pB*) [27].
5. Lipofectamine-LTX or any other transfection reagent specific to the cell line being used.
6. G418.

2.4 Flow Cytometry

1. FACSCalibur system (BD Biosciences) or any other appropriate flow cytometry equipment.
2. For cell sorting, BD FACS Aria II cell sorter (BD Biosciences) or any other appropriate cell sorter equipment.

2.5 Genomic DNA Isolation, Library Preparation, and Limited Next-Generation Sequencing

1. Lysis buffer: 100 mM NaCl, 20-mM Tris, pH 7.6, 10-mM EDTA, pH 8.0, 0.5 % sodium dodecyl sulphate and 0.5 mg/ml proteinase K.
2. 60 % volume-saturated NaCl.
3. Ethanol.
4. Personal genome machine (PGM) manually barcoded forward primer: P-*NNNN*AGTTGAACTAGTAATAAAGGATCC and PGM barcoded reverse primer: P-*NNNN*TGACATGTT GTATGACGGTGTG (*see* **Note 5**).
5. Ion Plus Fragment Library Kit (Life Technologies).
6. Ion PGM Template OT2 200 Kit (Life Technologies).
7. Ion PGM 200 Sequencing Kit (Life Technologies).
8. Ion 314 chip (Life Technologies).
9. PGM sequencing platform (Life Technologies).

2.6 Generation, Injection of Embryonic Stem Cells, Embryo Harvest, and Imaging for In Vivo Experiments

1. V6.5 embryonic stem cells (ESCs) derived from F1 hybrid strain (C57BL/6×129/Sv) [28].
2. ESC medium: DMEM, 15 % fetal bovine serum, 1000 U/ml LIF, 1 % β-mercaptoethanol, 1 % non-essential amino acids, 1 %l-glutamine, 0.5 % penicillin/streptomycin.
3. 1× phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4.
4. 100 mm culture dishes with feeder cells.
5. 2 N (3.5 days postcoitus) C57BL/6 blastocysts.
6. Pseudopregnant ICR recipient female mice—2.5 days postcoitus.
7. 4 % Paraformaldehyde.
8. 15 % and 30 % sucrose in 1× PBS.
9. OCT compound.
10. SuperFrost Plus slides.
11. Vectashield.
12. Confocal laser scanning microscope.

3 Methods

3.1 Construction of Donor Vectors

1. Generate the *att*B-flanked PCR products using One*Taq* DNA polymerase or any other DNA polymerase with proofreading activity.
2. For each BP recombination reaction between a given *att*B PCR product and donor vector, add the following components

to 1.5 ml microcentrifuge tubes at room temperature and mix gently with a pipette:

attB PCR product (150 ng)—1–7 μl (*see* **Notes 6–8**).

pDONR vector (150 ng/μl)—1 μl.

1× TE Buffer, pH 8.0—to 8 μl.

3. Quickly vortex the BP Clonase II enzyme mix, twice (2 s each time), and add 2 μl of BP Clonase II enzyme mix to each sample above. Mix well by vortexing briefly twice (2 s each time)
4. Incubate reactions at 25 °C for 1 h (*see* **Note 9**).
5. Add 1 μl of the Proteinase K solution to each reaction. Incubate for 10 min at 37 °C.
6. Transform 2 μl of the transformation reaction to TOP10 competent *E. coli* cells (the remaining can be stored at −20 °C) and plate one-fifth of the transformants on LB agar spectinomycin plates.
7. Incubate overnight at 37 °C.
8. The following day, screen colonies using M13 forward and reverse primers. For colony PCR (*see* **Note 10**), determine the number of colonies intended to be screened. Set up a 96-well plate with 100 μl/well of LB media containing spectinomycin (100 μg/ml) and a similar number of PCR reactions with M13 forward (−20) and reverse primers and *Taq* polymerase. Using a pipette tip pick one colony, dip it in the PCR reaction cocktail containing One*Taq* DNA polymerase, and then into the correspondingly labeled LB-containing well. Incubate the inoculated LB-containing plate at 37 °C.
9. Set up a thermal cycler with the following conditions:

   Initial denaturation at 94 °C for 30 s

   25 cycles (*see* **Note 11**) at 94 °C for 30 s, 47 °C for 30 s, 68 °C for "n" seconds, where $n = 60$ s/kb

   Final extension at 68 °C for 5 min
10. Resolve PCR products on a 1 % agarose gel. Colonies without any insert are characterized by a background band of ~350 base pairs (bp). If an insert of "n" bp is expected then a band at "350+n" bp will show up (Fig. 2).
11. Once a candidate insert has been identified, the corresponding inoculum can be used to seed miniprep culture, which can subsequently be sequence confirmed using the aforementioned M13 primers.

3.2 Construction of Expression Reporters

Complete expression reporters are generated via four part recombineering using the destination vector and the three donor plasmids—the *tRFP* entry clone, the donor plasmid containing the 3′ UTR/siRNA or miRNA sensor of interest, and the pool of donor

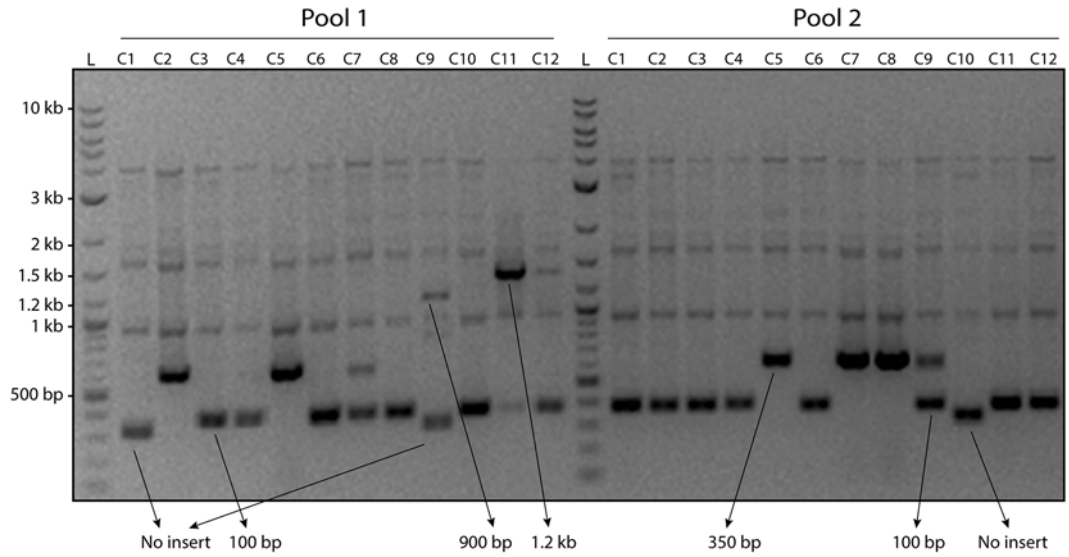

Fig. 2 Representative agarose gel (1 %) electrophoresis image of colony PCR products to identify positive clones in pooled BP recombination reaction. Colony PCR was performed with M13 primers (*see text*) to screen for positive clones from BP recombination reaction done in two separate pools. Clones without any insert result in a ~350 bp product. Subtracting ~350 bp from the other inserts gives the approximate length of the amplified 3′ UTR inserts and an indication of their identity. Performing colony PCR helps to pick the right size inserts for sequence confirmation in comparison to sequencing in bulk to get the right inserts. A similar strategy can be adapted for screening post LR recombination reaction, but using *UBC* forward and *PGK* reverse primers instead (*see text*). *L*, ladder; *C*, BP clone

plasmids containing the minimal polyadenylation signal and barcode.

1. Use the following formula to convert femtomoles (fmol) to nanograms (ng) of DNA:

   $$\text{ng} = [(\text{X fmol}) \times (\text{size of DNA in bp}) \times 660]/10^6$$

2. For each LR recombination reaction between an appropriate *att*B PCR product and donor vector (*see* **Note 12**), add the following components to 1.5 ml microcentrifuge tubes at room temperature and mix gently with a pipette:

   Entry clone (10 fmol each)—1–7 μl

   Destination *pBUTR* vector (20 fmol)—1 μl

   1× TE buffer, pH 8.0—to 8 μl

3. Quickly vortex the LR Clonase II Plus enzyme mix, twice (2 s each time), and add 2 μl of LR Clonase II enzyme mix to each sample above. Mix well by vortexing briefly twice (2 s each time).

4. Incubate reactions at 25 °C for 16 h (*see* **Note 13**).

5. Add 1 μl of the Proteinase K solution to each reaction. Incubate for 10 min at 37 °C.

6. Transform 2 μl of the transformation reaction to One Shot Mach1 T1R Competent *E. coli* cells (the remaining can be stored at −20 °C) and plate the entire transformants on LB agar ampicillin plus kanamycin plates.
7. Incubate overnight at 37 °C.
8. The following day, screen colonies using *UBC* forward and *PGK* reverse primers. For colony PCR, determine the number of colonies intended to be screened. Set up a 96-well plate with 100 μl/well of LB media containing ampicillin plus kanamycin (100 μg/ml each) and a similar number of PCR reactions with *UBC* forward and *PGK* reverse primers and One*Taq* DNA polymerase. Using a pipette tip pick one colony, dip it in the PCR reaction cocktail, and then into the correspondingly labeled LB-containing well. Once done, run PCR (use the same conditions as Subheading 3.1, **step 10**, except for annealing temperature of 49 °C). Incubate the inoculated LB-containing plate at 37 °C.
9. Resolve PCR products on an 1 % agarose gel. The right colonies can be identified based on the expected insert sizes of the 3′ UTR or siRNA/miRNA sensor elements.
10. Once an insert has been identified, the corresponding inoculum can be used to seed bacterial growth cultures, which would subsequently be sequence confirmed using the *UBC* forward and *PGK* reverse primers described above (*see* **Note 14**).

3.3 Transfection of Cells and Generation of Stable Clones

1. Transfect cells with plasmids containing transposase (*pCMV-HA-m7pB*) and transposon (respective *pBUTR* vector) at a ratio of 1:2 using appropriate transfection reagent (*see* **Note 15**). The *pBUTR* can be used for a wide spectrum of cell types (*see* **Note 16**). For MCF10A cells, seed 4×10^4 cells into each of the desired number of wells in a 24-well plate. Twenty-four hours after cell seeding, transfect the cells in each well with 333.3 ng of *pBUTR-wild-type-ZEB2* or *pBUTR-mutant-ZEB2* along with 166.7 ng of *pCMV-HA-m7pB*.
2. Forty-eight hours after transfection, split cells 1:10 and select with G418 (1000 μg/ml for MCF10A) for approximately 2 weeks (*see* **Note 17**).

3.4 Flow Cytometric Analysis of Reporter Expression in Arrayed Format

1. Following G418 selection, split each stably transduced cell line in replica plates, and then treat one or more replicates with the experimental stimulus while leaving another replicate plate untreated as a control. For MCF10A cells, treat with TGF-β or vehicle for 72 h (*see* **Note 18**, and Fig. 3a).
2. Perform multicolor flow cytometry to assess the expression of the turboGFP (tGFP) (excitation/emission max = 482/502 nm) and tRFP (excitation/emission max = 553/574 nm) under the different experimental conditions.

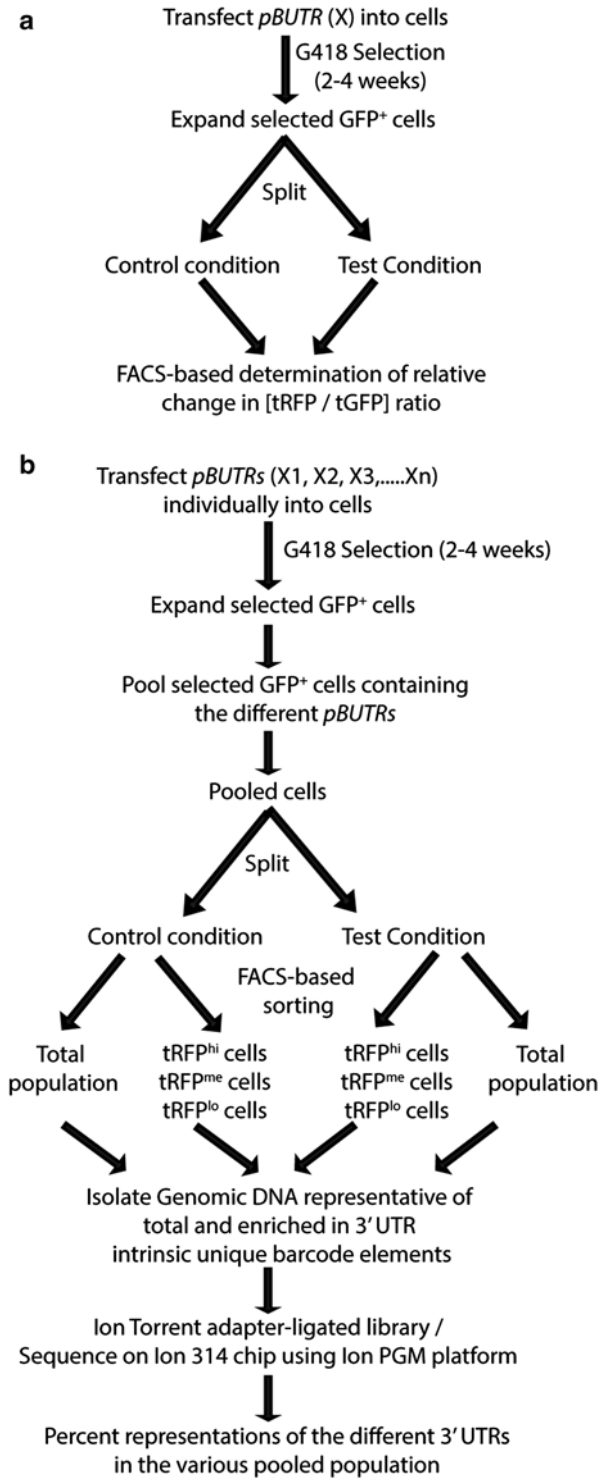

Fig. 3 Schematic of work-flow for *pBUTR* vector-mediated high-dimensionality screening and validation applications in arrayed (**a**) or pooled format (**b**). One caveat associated with DNA transposon-based screening approaches relative to a retro- or lentiviral approach is that stable transfection of cells in bulk with a pool of vectors is not straightforward. For this reason, initial transfection and selection should be performed in an arrayed format. The inclusion of unique barcode elements with the minimal polyadenylation signal will allow analysis of enrichment or depletion within pooled cell populations via flow cytometry and cell sorting and limited next-generation sequencing

3. Determine the ratio of tRFP and tGFP expression as assessed via median fluorescence intensity (MFI) and calculate fold changes as follows:

   Fold change = $\log_2$ $(\text{tRFP}/\text{tGFP})_{C1}/(\text{tRFP}/\text{tGFP})_{C2}$, where C1 and C2 are two different experimental conditions.

4. A positive fold change will indicate 3′ UTR-mediated post-transcriptional induction or up-regulation of gene expression. Conversely, a negative fold change will indicate posttranscriptional repression or downregulation of gene expression (*see* **Notes 19** and **20**).

Treatment of MCF10A cells with TGF-β will result in decreased miR-200b and E-cadherin protein expression. In cells stably transfected with the wild-type *ZEB2* reporters, these decreased levels will coincide with marked increases in tRFP fluorescence. The levels of tRFP fluorescence, as assessed via median fluorescent intensity (MFI) in the TGF-β treated MCF10A cells will be similar to those observed in untreated MCF10A cells transfected with mutant *ZEB2* reporters (*see* **Note 21**).

3.5 Flow Cytometric Analysis of Reporter Expression in Pooled Format

1. The *pBUTR* vectors containing the 3′-UTR elements of interest are individually transfected in arrayed format and selected with G418 exactly as described in Subheading 3.3 (*see* **Note 22**).
2. Following selection for approximately 2–4 weeks, the individual transfectants expressing the different *pBUTRs* are pooled (*see* **Note 23**, and Fig. 3b).
3. The pooled stable transfectant lines are split to replica plates and treated according to experimental design.
4. Cells from the different experimental conditions are sorted using multicolor flow cytometry based on tRFP expression (*see* **Notes 24** and **25,** and Fig. 4).
5. Isolate genomic DNA from the different pools of sorted tRFP+ positive cells using overnight proteinase K digestion at 55 °C before salting out with 60 % volume saturated NaCl and precipitating with ethanol.
6. Use distinctly barcoded PGM PCR primer pairs to amplify 3′ UTR-correlated barcode elements from each of the sorted populations.
7. Use these barcoded elements to template an Ion Torrent adapter-ligated library using Life Technology's Ion Plus Fragment Library Kit protocol (#4471252, Revision 3.0). Perform sample emulsion PCR using the Ion PGM Template OT2 200 Kit (#4480974, Revision 5.0) following Life Technology's instructions. Prepare the samples for sequencing using the Ion PGM 200 Sequencing Kit (#4474004,

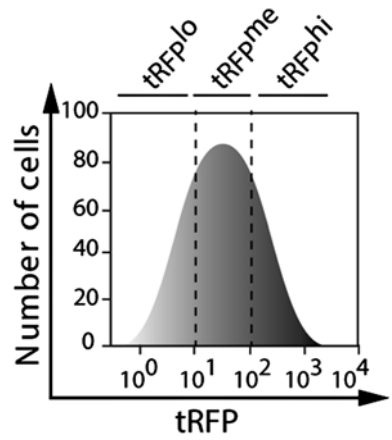

Fig. 4 Recommended subdivision of population prior to screening for more sensitivity in pooled approaches. Shown is a schematized example of the range of basal tRFP expression of a pool of *pBUTRs* in a given physiological context. The original pool may be sorted into two or more subpools (*vertical dotted lines*) on the basis of baseline tRFP expression (e.g., $tRFP^{lo}$, $tRFP^{med}$, $tRFP^{hi}$) prior to the experiment. Each subpool may then be individually treated and sorted. Please also refer to **Note 25**. *tRFP*, turbo-red fluorescent protein; *lo*, low; *me*, medium; *hi*, high

Revision C). Load the completed samples on an Ion 314 chip and sequence on the PGM platform.

8. Process the data from the PGM runs initially using the bam-2fastq [29] to generate the fastq files and custom Perl scripts to bin based on barcodes and trim adapter sequences. Determine the percent representation of the different barcodes in the indicated populations.
9. Enriched and depleted barcodes will reveal identity of gene products that are being regulated at the post-transcriptional level by their 3′ UTR elements under different experimental conditions.

3.6 Using pBUTR for In Vivo Monitoring of 3′ UTR-Mediated PTGR

1. On the day of electroporation, trypsinize, count and aliquot 5×10^6 V6.5 ESCs.
2. Spin the tube containing the aliquoted cells at $190 \times g$ for 3 min. Aspirate as much media off as possible, wash cell pellet with 10 ml 1× PBS, and spin again for 3 min at $190 \times g$. Aspirate once again.
3. Add 1 μg each of *pBUTR* vector and *pCMV-HA-m7pB* transposase to the cell pellet.
4. Add 700 μl 1× PBS to the pellet. Suspend cells and DNA by pipetting multiple times and transfer to 0.4 cm gap electroporation cuvette. Replace cap on cuvette (*see* **Note 26**).

5. Place cuvette in GenePulser shockpod and electroporate the cells at 240v and 500uF (*see* **Note 27**).
6. Post-electroporation, let the cuvette sit inside a laminar air flow hood for 10 min at room temperature.
7. Add 1 ml of ESC media to the cuvette and mix. Transfer cells to a 15 ml conical tube in sterile condition (*see* **Note 28**).
8. Rinse cuvette with 1 ml of ESC media and add to the cells in the conical media. Resuspend cells ensuring there are no clumps.
9. Plate 10 % of electroporated cells onto 100 mm culture dishes with a feeder layer by gently dripping the cells over the feeder layer.
10. Tip dishes in "X" pattern (do not swirl cells to the periphery of the plate).
11. At least 20 h post-electroporation, start selection with G418 (300 μg/ml) for 8 days, verify resulting ESC colonies for tGFP and tRFP expression using a microscope, pick the colonies, expand, and make freezer stocks.
12. Inject ES cell clones into 2 N (3.5 days postcoitus) C57BL/6 blastocysts and subsequently transfer to the uterine horns of 2.5 days postcoitus pseudopregnant ICR recipient female mice.
13. Sacrifice pregnant females by carbon dioxide asphyxiation on appropriate postcoitus day based on what developmental phase is being studied (*see* **Note 29**).
14. Dissect out embryos and fix in 4 % paraformaldehyde (PFA) for 1 h, before incubation in 15 % and 30 % sucrose (each for 16 h), and finally embed in OCT compound.
15. Cut 5 μm sections including desired physiological structure and mount on SuperFrost Plus slides using Vectashield.
16. Obtain images documenting domains of tRFP and tGFP expression using a confocal laser scanning microscope.
17. TurboGFP expression would be constitutively observed in all areas where the *pBUTR* has been internalized. On the other hand, the tRFP expression will be reliant on a particular 3′ UTR's ability to confer correct temperospatial expression of the gene product during murine development.

4 Notes

1. For generating the siRNA or miRNA sensors, the *att*B2r and *att*B4 flanked siRNA/sensors sequence may be commercially synthesized. For example, for a "2×" *CXCR4* siRNA sensor [30] the *att*B2r-*CXCR4*-*att*B4 sequence is—5′ –GGGG

A C C C A G C T T T C T T G T A C A A A G T G G T A A G T T T T C A C T C C A G C T A A C A C C G G A T C G G C A T A A G T T T T C A C T C C A G C T A A C A C C G G C A C C C A A C T T T T C T A T A C A A A GTTGTCCCC—3′ (the underlined portion is the *CXCR4* sensor sequence). This commercially synthesized oligonucleotide may then be amplified with oligonucleotides corresponding only to the *attB2r* and *attB4* sites, rather than chimeric oligonucleotides as described.

2. For generating an *att*B-flanked entire 3′ UTR, design the forward primer with the 5′ end corresponding to the base immediately after the stop codon of the coding sequence and the 5′ end of the reverse primer corresponding to the nucleotide immediately preceding the poly (A) signal. The use of a synthetic polyadenylation signal for all clones to be analyzed removes any influence of the native polyadenylation signal (e.g., efficiency) on gene expression, which may confound analysis of 3′ UTR-mediated effects.
3. The composition of our own barcodes, generated via mixed nucleotide synthesis, was informed by the average nucleotide composition of the 24 base pairs following the G/U-rich region of native polyadenylation sequences in the human genome. The inclusion of unique barcode elements with the minimal polyadenylation signal was made to allow analyses within pooled cell populations via flow cytometry, cell sorting, and limited next-generation sequencing analysis.
4. The BP Clonase II and LR Clonase II Plus enzyme mixes should be kept at –20 °C until immediately before use; however, the Proteinase K solution can be thawed and kept on ice until use.
5. The number of different 'NNNN' combinations to be used will depend on the experimental conditions. For example, if only a control and experimental conditions are being compared then two variants of NNNN like ACTG and AGTC will be used. On the other hand, if a time course experiment is being done for 0, 24, and 48 h, then three variants of NNNN like ACTG, TGAC, and AGTC may be used. These primer pairs should be phosphorylated at the 5′ end.
6. Anywhere between 15 and 150 ng of the *att*B PCR product can be used for the BP reaction.
7. Set up a BP reaction with no *att*B PCR product as a negative control.
8. A major advantage of the Gateway® system is the potential for scalability. Multiple donor vectors containing 3′ UTR elements to be assessed can be generated simultaneously if the *att*B2r and *att*B4 flanked PCR products corresponding to these 3′

UTRs are pooled for the BP reaction. For pooling of large groups of UTRs it is recommended to generate "subpools" such that individual inserts may be easily discerned following colony PCR screening and gel electrophoresis. For example if there are 3′ UTRs of length 150, 175, 200, 250, 500, 750, 1200, 1400 bp—then we would recommend two subpools, with one containing 3′ UTRs of length 150, 200, 500, and 1200 bp and the other containing the 175, 250, 750, and 1400 bp length 3′ UTRs. This facilitates identification of individual inserts during visualization after gel electrophoresis.

9. Normally a 1-h incubation yields a sufficient number of donor vectors. However, the length of the recombination reaction can be extended up to a maximum of 18 h. For PCR products ≥5 kb, longer incubations will increase the yield of colonies and are recommended. Normally, an overnight incubation typically yields five to ten times more colonies than 1-h incubation.
10. If just one donor vector is being constructed then the colonies can just be grown up for miniprep and subsequently sequenced. The colony PCR is especially beneficial when pooled BP reactions are being done since this precludes the need to sequence a large number of colonies to get the desired donor vectors.
11. Normally 25 cycles of PCR is enough to view products on an agarose gel. The precise conditions for PCR will be informed by the choice of polymerase mix and the thermal cycler used.
12. The expression constructs can be generated through pooled LR recombination reaction in a manner analogous to that described in *see* **Note 8**. Again, we recommend a subpooling strategy based on the size of the 3′ UTR elements of interest (*see* **Note 8**) such that the positive clones can be easily identified through colony PCR.
13. The reactions can be incubated up to 24 h at room temperature.
14. The *pBUTR* reporters should be prepped with Endotoxin-free miniprep or maxiprep kits, depending on the number of projected downstream transfections.
15. The total amount of DNA transfected depends on the cell numbers and the specific transfection reagent being used.
16. Originally isolated from the genome of the cabbage looper moth *Trichoplusia ni* [31], the *piggyBac* transposon has distinct advantages. It has a large cargo size [31], and is highly active in many cell types [32, 33]. In addition, it has been shown to effect long-term expression in mammalian cells in vivo [34].

17. Perform a kill curve with G418 for the particular cell line being used, then use the lowest concentration of G418 that effectively kills untransfected cells. Normally you get stable colonies of cells within 2 weeks and very distinct isolated colonies after 4 weeks of selection with G418. Successful selection can be confirmed by observing the cells under a fluorescence microscope and determining the approximate percentage of $tGFP^+$ cells.
18. Posttreatment with TGF-β for 72 h, MCF10A cells switch from polarized, tightly packed discoid epithelial cells to highly motile fibroblastic or mesenchymal phenotype, characteristics of distinct morphological changes associated with EMT [26], a reduction in E-cadherin protein expression concomitant with an induction of the mesenchymal cell marker N-cadherin [18]. These parameters can be used to verify that the answer obtained in the *pBUTR* experiment corroborates with the expected landmarks of a phenomenon.
19. The observed changes in tRFP expression can be further validated by appropriately using miRNA mimics or antagomirs in the case of siRNA/miRNA sensors or miRNA-mediated regulation or siRNA/ectopic overexpression in case of RNA binding proteins.
20. The relative reporter expression within this system does not differentiate between mechanisms impacting mRNA stability or translational repression, which will require additional downstream experimentation.
21. Of note, even though endogenous promoters are used in the *pBUTR* vector, it is necessary to include appropriate control reporters with minimal or otherwise defined 3′ UTR elements to offset effect of promoter activity, if any.
22. A drawback of DNA transposon-based approaches is that there is some risk in transfecting pooled reporters into a population of cells. In contrast to viral vector systems, where low multiplicities of infection can be used to ensure a single integrant per cell, DNA transposons necessitate the use of electroporation or cationic lipid-based delivery methods. Since both of the latter methods will deliver multiple vectors from a pool into a given cell, there is a very high risk of confounding results in any reporter-based screen. We thus strongly suggest that individual cell lines be generated in arrayed format and then pooled for screening approaches.
23. Ideally, equal numbers of cells from each stably transduced line are mixed together. However, since a comparison of relative representation within control and experimental populations will be assessed, this is not essential in high-dimensionality screens.

24. An advantage of flow cytometry based screening is that additional fluorophores may be simultaneously used as experimental controls, e.g., decreased surface expression of E-cadherin and increased surface expression of N-cadherin in the context of EMT.
25. As a general rule, one would collect the 10 % of cells expressing the highest level of tRFP in each condition and compare these populations. However, this strategy may miss several posttranscriptional regulatory events. Each 3′ UTR has its own baseline level of expression, which from our experience may vary over an order of magnitude from other 3′ UTRs in the population. For example, consider that the basal tRFP fluorescence intensity of a pooled population ranges from 10 to 100 (arbitrary units). An individual reporter may have a fluorescence intensity of 10 in the control state and 80 in the experimental state—an impressive eightfold induction. However, since one is merely collecting the top 10 % of events in each population (fluorescence intensities of 90–100) this induction would be missed. To increase the sensitivity in a screening experiment, the original pool may be sorted into two or more subpools on the basis of baseline reporter expression (e.g., $tRFP^{lo}$, $tRFP^{med}$, $tRFP^{hi}$) prior to the experiment. Each subpool may then be individually treated and sorted.
26. While loading the cuvette, be careful not to touch the sides, especially the metallic surface.
27. Confirm that the Time Constant from the GenePulser was between 7.0 and 8.0 during electroporation.
28. The mixing is done best with Pasteur pipettes.
29. Although we describe a transient transgenesis approach, depending on the depth and breadth of the planned analysis, a better strategy may be to let the fetuses come to term and screen pups for tGFP expression upon birth [35]. $TGFP^{+}$ pups may then be used as founders for a line of reporter mice that may be used to extensively characterize 3′ UTR-mediated gene regulation throughout embryogenesis and adulthood.

References

1. Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470
2. Cloonan N, Forrest AR, Kolle G et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 5:613–619
3. Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628
4. Chen G, Gharib TG, Huang CC et al (2002) Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics 1:304–313
5. Vogel C, Abreu RS, Ko D et al (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol Syst Biol 6:400

6. Schwanhausser B, Busse D, Li N et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–342
7. Jansen RP (2001) mRNA localization: message on the move. Nat Rev Mol Cell Biol 2:247–256
8. de Moor CH, Meijer H, Lissenden S (2005) Mechanisms of translational control by the 3′ UTR in development and differentiation. Semin Cell Dev Biol 16:49–58
9. Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. Nat Rev Mol Cell Biol 8:113–126
10. Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543
11. Merritt C, Rasoloson D, Ko D et al (2008) 3′ UTRs are the primary regulators of gene expression in the C. elegans germline. Curr Biol 18:1476–1482
12. Matoulkova E, Michalova E, Vojtesek B et al (2012) The role of the 3′ untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol 9:563–576
13. Kornblihtt AR, Schor IE, Allo M et al (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14:153–165
14. Elkon R, Ugalde AP, Agami R (2013) Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet 14:496–506
15. Conne B, Stutz A, Vassalli JD (2000) The 3′ untranslated region of messenger RNA: A molecular 'hotspot' for pathology? Nat Med 6:637–641
16. Shibayama A, Cook EH Jr, Feng J et al (2004) MECP2 structural and 3′-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. Am J Med Genet B Neuropsychiatr Genet 128B:50–53
17. Chatterjee S, Pal JK (2009) Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. Biol Cell 101:251–262
18. Chaudhury A, Kongchan N, Gengler JP et al (2014) A piggyBac-based reporter system for scalable in vitro and in vivo analysis of 3′ untranslated region-mediated gene regulation. Nucleic Acids Res 42, e86
19. Nakano K, Ando T, Yamagishi M et al (2013) Viral interference with host mRNA surveillance, the nonsense-mediated mRNA decay (NMD) pathway, through a new function of HTLV-1 Rex: implications for retroviral replication. Microbes Infect 15:491–505
20. Zufferey R, Donello JE, Trono D et al (1999) Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J Virol 73:2886–2892
21. Jahner D, Stuhlmann H, Stewart CL et al (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature 298:623–628
22. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–340
23. Jones S, Peng PD, Yang S et al (2009) Lentiviral vector design for optimal T cell receptor gene expression in the transduction of peripheral blood lymphocytes and tumor-infiltrating lymphocytes. Hum Gene Ther 20:630–640
24. Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome Res 10:1788–1795
25. Xia H, Mao Q, Davidson BL (2001) The HIV Tat protein transduction domain improves the biodistribution of beta-glucuronidase expressed from recombinant viral vectors. Nat Biotechnol 19:640–644
26. Gregory PA, Bert AG, Paterson EL et al (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 10:593–601
27. Yusa K, Zhou L, Li MA et al (2011) A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A 108:1531–1536
28. Eggan K, Akutsu H, Loring J et al (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. Proc Natl Acad Sci U S A 98:6209–6214
29. http://www.hudsonalpha.org/gsl/information/software/bam2fastq.
30. Doench JG, Petersen CP, Sharp PA (2003) siRNAs can function as miRNAs. Genes Dev 17:438–442
31. Cary LC, Goebel M, Corsaro BG et al (1989) Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. Virology 172:156–169
32. Li MA, Turner DJ, Ning Z et al (2011) Mobilization of giant piggyBac transposons in the mouse genome. Nucleic Acids Res 39, e148
33. Ding S, Wu X, Li G et al (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell 122:473–478

34. Wu SC, Meir YJ, Coates CJ et al (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci U S A 103:15008–15013

35. Nakanishi H, Higuchi Y, Kawakami S et al (2010) piggyBac transposon-mediated long-term gene expression in mice. Mol Ther 18:707–714

Part III

RBP Interactomics

Chapter 8

Comprehensive Identification of RNA-Binding Proteins by RNA Interactome Capture

Alfredo Castello, Rastislav Horos, Claudia Strein, Bernd Fischer, Katrin Eichelbaum, Lars M. Steinmetz, Jeroen Krijgsveld, and Matthias W. Hentze

Abstract

RNA associates with RNA-binding proteins (RBPs) from synthesis to decay, forming dynamic ribonucleoproteins (RNPs). In spite of the preeminent role of RBPs regulating RNA fate, the scope of cellular RBPs has remained largely unknown. We have recently developed a novel and comprehensive method to identify the repertoire of active RBPs of cultured cells, called RNA interactome capture. Using in vivo UV cross-linking on cultured cells, proteins are covalently bound to RNA if the contact between the two is direct ("zero distance"). Protein-RNA complexes are purified by poly(A) tail-dependent oligo(dT) capture and analyzed by quantitative mass spectrometry. Because UV irradiation is applied to living cells and purification is performed using highly stringent washes, RNA interactome capture identifies physiologic and direct protein-RNA interactions. Applied to HeLa cells, this protocol revealed the near-complete repertoire of RBPs, including hundreds of novel RNA binders. Apart from its RBP discovery capacity, quantitative and comparative RNA interactome capture can also be used to study the responses of the RBP repertoire to different physiological cues and processes, including metabolic stress, differentiation, development, or the response to drugs.

Key words RNA, RNA-binding protein, Proteomics, Proteome, RNA interactome capture, Posttranscriptional regulation, Gene expression

1 Introduction

In the last decade, in vitro and in silico approaches have been developed in order to determine the complete repertoire of RNA-binding proteins (RBPs), referred to here as the RNA interactome. While in vitro approaches served to identify dozens of novel RBPs [1, 2], the abundant negative charges of the sugar-phosphate backbone of the RNA have the capacity to mediate unspecific binding of basic proteins in vitro and, therefore, extensive validation is required. Biocomputational methods can recognize proteins bearing classical RNA-binding domains (RBDs) such as the RNA

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_8,

recognition motif (RRM) and K-homology domain (KH) [3]. However, their capacity to identify unorthodox RBDs is very limited. To address these limitations we and others developed RNA interactome capture, which combines UV cross-linking and oligo(dT) capture to pull down proteins bound to polyadenylated RNA in living cells. Applied to HeLa [4] and HEK293 cells [5], RNA interactome capture determined the first near-complete RNA interactomes of a human cell line. More recently, this approach has also been applied successfully to mouse embryonic stem cells [6] and *Saccharomyces cerevisiae* [7].

In a first step, protein-RNA interactions are "fixed" applying two different approaches: (a) Irradiation with ultraviolet (UV) light at 254 nm of cell monolayers induces short-lived radicals at the nucleotide base that can attack amino acids in close proximity forming covalent bonds. (b) The photoactivatable-ribonucleoside-enhanced cross-linking (PAR-CL) protocol employs 4-thiouridine (4SU) [8], which is taken up by cells and incorporated into nascent RNAs. Protein-RNA cross-linking is achieved by irradiation at 365 nm. Following UV cross-linking by either approach, lysis under denaturing conditions, and homogenization, polyadenylated RNAs and their covalently bound proteins are isolated with oligo(dT) magnetic beads through highly stringent washes. RBPs bound to polyadenylated RNA are then released by RNase treatment and identified by proteomics.

RNA interactome capture has notable advantages over previous RBP identification methods: (1) As UV irradiation is applied to cell monolayers, protein-RNA interactions discovered have occurred within the native context and without overexpression. (2) Because the free radicals are induced at the nucleotide base, UV irradiation promotes exclusively protein-RNA and not protein-protein cross-links [9, 10]. (3) The short-lived nature of the free radicals (nanosecond range) limits covalent bond formation to amino acids at "zero distance" (~2 Å) [9]. (4) Due to the stability of nucleic acid hybrids (i.e., poly(A) tails binding to oligo(dT) magnetic beads) in the presence of high-salt and chaotropic detergents, very stringent washing steps can be applied to remove all non-cross-linked polypeptides. On the other hand, RNA interactome capture will fail to detect RBPs when they are (1) not bound to polyadenylated RNAs, (2) not expressed in the cell type under study, (3) not active in RNA binding under the experimental conditions, or (4) not cross-linked efficiently by UV irradiation.

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents. Store buffers at 4 °C and samples at −70–80 °C (unless indicated otherwise). All buffers should be filtered and autoclaved

before usage. For buffers containing LiDS or DTT, autoclaving should be performed prior to the addition of these heat-sensitive components.

2.1 Buffers

1 Lysis buffer: 20 mM Tris–HCl pH 7.5, 500 mM LiCl, 0.5 % LiDS (wt/v, stock 10 %), 1 mM EDTA, 5 mM DTT.

2 Buffer 1: 20 mM pH 7.5 Tris–HCl, 500 mM LiCl, 0.1 % LiDS (wt/v), 1 mM EDTA, 5 mM DTT.

3 Buffer 2: 20 mM pH 7.5 Tris–HCl, 500 mM LiCl, 1 mM EDTA, 5 mM DTT.

4 Buffer 3: 20 mM pH 7.5 Tris–HCl, 200 mM LiCl, 1 mM EDTA, 5 mM DTT.

5 Elution buffer: 20 mM pH 7.5 Tris–HCl, 1 mM EDTA.

6 Buffer 4: 50 mM NaCl.

7 10× RNase buffer: 50 mM pH 7.5 Tris–HCl, 1.5 M NaCl, 5 mM DTT.

8 5× Proteinase K buffer: 50 mM pH 7.5 Tris–HCl, 750 mM NaCl, 1 % SDS, 50 mM EDTA, 2.5 mM DTT, 25 mM $CaCl_2$.

9 Phosphate-buffered saline (PBS).

2.2 Reagents

1. Oligo (dT_{25}) magnetic beads (New England Biolabs, S1419S).
2. 4-Thiouridine (4SU, Sigma).
3. Lithium dodecyl sulfate (LiDS).
4. Lithium chloride (LiCl).
5. DTT.
6. Amicon Ultra® Centrifugal Filters (50 ml, 10 KDa cutoff, Millipore UFC901024).
7. Ribonuclease T1 from *Aspergillus oryzae* (RNase T1).
8. Ribonuclease A from bovine pancreas (RNase A).
9. RNeasy kit (Qiagen).

2.3 Equipment

1. Humidified 37 °C, 5 % CO_2 incubator.
2. Microbiological cabinet class 2.
3. 15 cm dishes or 500 mm^2 square dishes.
4. Cross-linking devices: cCL = 254 nm bulbs; PAR-CL= 365 nm bulbs. Spectrolinker UV Cross-linkers (Spectroline).
5. Needle (27G, 3/4-inch; no. 20, 0.4 mm × 19 mm).
6. Sterile syringe (5 ml).
7. 50 ml Magnetic separation rack and 12-tube (2 ml) magnetic separation rack.
8. Refrigerated bench-top centrifuge.

3 Methods

Experimental design: Include a non-irradiated (noCL) sample to control the signal obtained for UV-irradiated samples for background noise derived from the purification with oligo(dT) magnetic beads [11]. In the same vain, include 4SU-treated, non-irradiated cells (4SU noCL) as a control for PAR-CL. The overview of the cross-linking and purification protocol is schematized in Fig. 1a–c.

1. For conventional cross-linking (cCL), seed cells in 5×500 cm^2 dishes (1500 cm^2 of total growth area) with normal medium (e.g., DMEM and 5 % fetal calf serum, FCS) to reach 80–90 % confluence after overnight incubation (~1.9×10^7 cells per dish; *see* **Note 1**) (Fig. 1a). For PAR-CL, follow the same seeding protocol, but supplement the medium with 100 μM 4SU (*see* **Note 2**).
2. After overnight incubation, wash cells twice with 30 ml of PBS (room temperature) until the PBS remains colorless.
3. Remove the PBS and place the culture dishes without their lids on ice at ~15 cm from the UV source. Irradiate with 150 mJ/cm^2 at 254 nm UV light for cCL or at 365 nm UV light for PAR-CL (*see* **Note 3**). After irradiation, add 15 ml of ice-cold PBS per dish and keep the already irradiated dishes at 4 °C while processing the rest of the dishes.
4. Scrape the cells into the PBS added in **step 3** with a rubber policeman and centrifuge at $400 \times g$ for 3 min at 4 °C (*see* **Note 4**). Remove and discard the supernatant.
5. Add 10 ml of ice-cold lysis buffer, resuspend the pellet pipetting up and down, and supplement the lysate with 30 additional ml of lysis buffer. Mix the lysate by inverting the 50 ml tube (*see* **Note 5**).
6. Pass the lysate through a 5 ml syringe with a narrow needle (gauge 0.4 mm diameter) to homogenize. Repeat the process two additional times until the viscosity of the lysate is significantly reduced (*see* **Notes 6** and **7**). Keep the sample in ice while processing the other samples and the oligo(dT) beads are equilibrated (see below).
7. Equilibrate 2 ml of oligo(dT)$_{25}$ magnetic beads per tube by washing three times with 5× volumes of lysis buffer. Resuspend the bead pellet in 2 ml of lysis buffer (*see* **Note 5**).
8. Add resulting bead suspension (2 ml) to the sample and incubate for 1 h at 4 °C with gentle rotation (*see* **Note 5**).
9. Place the tubes on a magnet at 4 °C and wait until the beads are collected (this can take up to 30 min). Recover the supernatant and store it in a new tube at 4 °C for the following two cycles of oligo(dT) capture (see below).

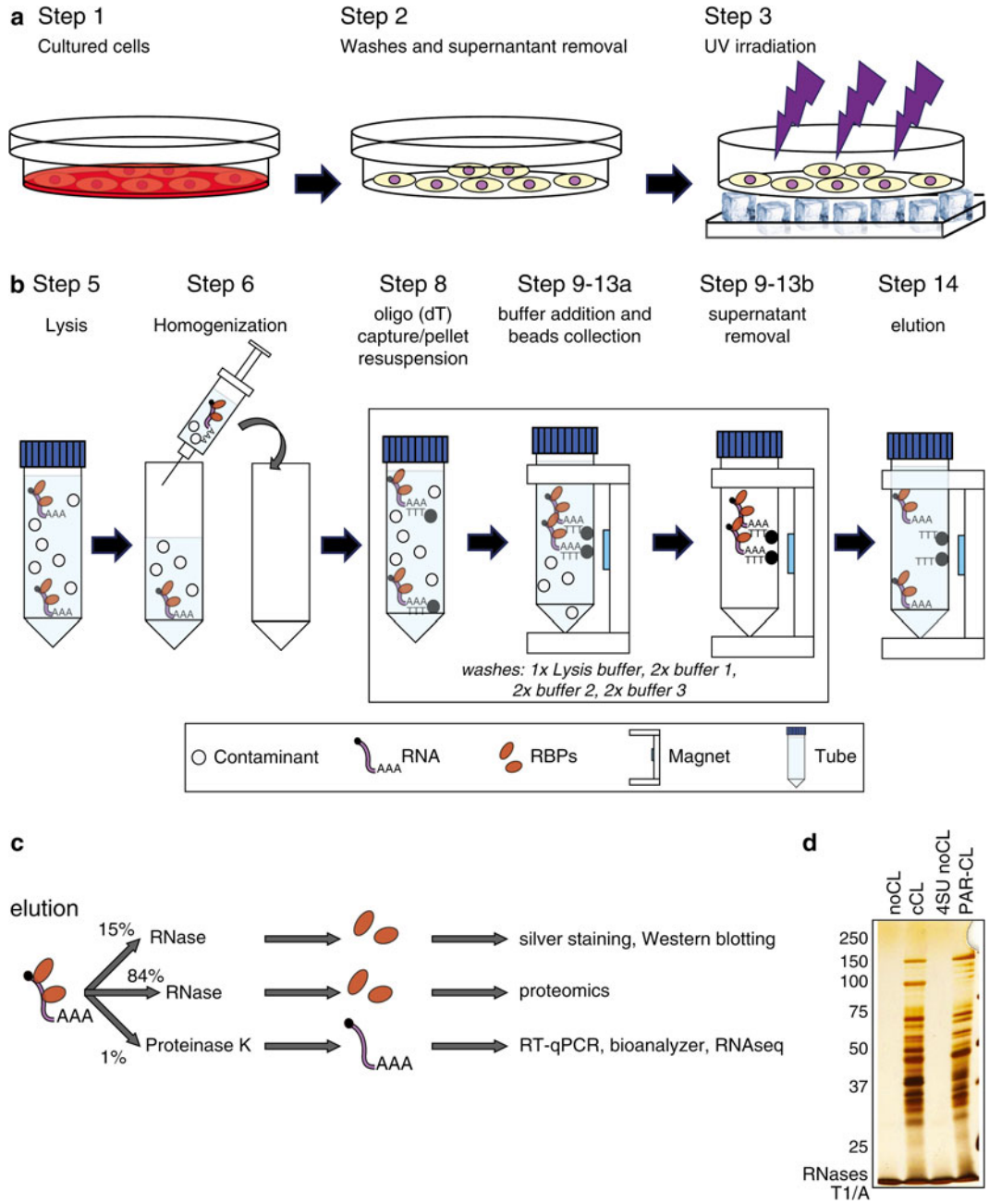

Fig. 1 RNA interactome capture workflow: Schematic representation of in vivo UV cross-linking (**a**), oligo (dT) capture (**b**), and downstream processing of the eluates (**c**). After elution, samples are treated with either proteinase K for RNA quality controls or RNases for protein quality controls and mass spectrometry. (**d**) Representative silver staining of RNase-treated eluates

10. Resuspend the bead pellet in 35 ml ice-cold lysis buffer. Incubate for 5 min at 4 °C with gentle rotation and pellet the beads with the magnet. Discard the supernatant.
11. Add 35 ml of ice-cold buffer 1, resuspend the beads, and incubate for 5 min at 4 °C with gentle rotation. Pellet the beads with the magnet and discard the supernatant (*see* **Notes 8** and **9**). Repeat this step once.
12. Add 35 ml of ice-cold buffer 2 and resuspend the beads. Mix by inverting the sample ten times. Pellet the beads with the magnet and discard the supernatant (*see* **Notes 8** and **9**). Repeat this step once.
13. Add 35 ml of ice-cold buffer 3, and resuspend the beads. Mix by inverting the sample ten times. Pellet the beads with the magnet and discard the supernatant. Repeat this step once.
14. Resuspend the bead pellet in 500 μl of elution buffer. Transfer the sample to a sterile 1.5 ml tube and elute the RNA-protein complexes by incubating at 55 °C for 3 min. Collect the beads in a magnet and transfer the supernatant to a new tube. Pellet any residual beads again in the magnet and collect the supernatant and transfer it to a new sterile tube (*see* **Note 10**). Determine the resulting RNA content using a Nanodrop device.
15. Recycle the beads following the manufacturer's recommendation. Add recycled beads to the lysate stored at 4 °C (*see* **step 9**) and repeat the isolation (from **steps 8** to **14**) twice (three isolation cycles in total).
16. Pool the eluates from the three successive oligo (dT) capture cycles (final volume 1.5 ml).
17. Take 20 μl of the pooled eluate from **step 16**. Add 5 μl of 5× proteinase K buffer and 1 μg of proteinase K, incubate for 30 min at 37 °C and 30 min at 50 °C. Next, isolate the RNA with RNeasy kit (Qiagen) or trizol (Invitrogen). Use purified RNA for RNA quality control analyses (e.g., RT-qPCR using primers against mRNAs and rRNAs, bioanalyzer, or RNAseq) [4, 11].
18. Take the rest of the elution and add 150 μl of 10× RNase buffer, and ~50–100 U of RNase T1 and RNase A. Incubate at 37 °C for 1 h followed by 15 min at 55 °C.
19. Transfer the eluate into an Amicon Ultra 10^{-3} KDa cutoff (*see* **Note 11**). Top up the filter device with buffer 4 and centrifuge at 4000 × *g* for 45 min at 4 °C.
20. Discard the flow through and top up the filter device again with buffer 4. Centrifuge at 4000 × *g* for 45 min at 4 °C.
21. Recover the sample from the filter unit in about 200 μl.
22. Use 30 μl of the sample for protein quality analyses (e.g., silver staining in Fig. 1d) [4, 11].

23. Once quality controls are performed satisfactorily, the rest of the sample can be analyzed by quantitative mass spectrometry. RNA interactome capture is compatible with all the state-of-the-art quantitative proteomic approaches, including label-free quantification [4], SILAC [5], dimethyl labeling [6], and isobaric labeling (tandem mass tags, TMT).

4 Applications

RNA interactome capture has been used to determine the RNA-bound proteome of HeLa [4], HEK293 [5], mouse embryonic stem cells [6], and *S. cerevisiae* [7]; it can readily be applied to other cell lines and primary cells, and likely be adapted to organisms. Finally, RNA interactome capture can be used in a quantitative and comparative way to explore the plasticity of mRNA interactomes in response to different physiological conditions and biological cues. Moreover, RNA interactome capture has been used to study the RNA-binding capacity of a given protein in vivo. In brief, the protein of interest is fused to eGFP and expressed in cultured cells. Upon UV cross-linking and oligo (dT) capture, eGFP signal is measured in a plate reader and used as a proxy for RNA binding. This protocol requires 1/5 of the cells, buffer volume, and beads indicated above [12].

5 Notes

1. The cell number indicated in the protocol refers to HeLa cells and this may vary between cell lines due to differences in cell volume and RNA content per cell. A successful large scale RNA interactome capture experiment will yield ~100–300 μg of RNA upon oligo (dT) pull down. Knowing the amount of RNA isolated from a defined number of cells, it is possible to calculate the quantity of cells required to capture ~100–300 μg of RNA.
2. We observed that 100 μM is the optimal concentration of 4SU for most of the cell lines tested. Nevertheless, the 4SU dose may require optimization for certain cell lines.
3. Irradiation with 150 mJ/cm^2 of 254 nm UV light yields a relatively high UV cross-linking efficiency while keeping the RNA intact. In our hands, this dose is optimal for most adherent cell types. However, we have found few remarkable exceptions, suggesting that UV dosage may require optimization when working with different cell lines in order to maximize protein yield after oligo (dT) capture.
4. Some cells are sensitive to scraping. In these cases, we recommend to perform direct on-plate lysis to avoid loss of material.

Add the lysis buffer directly onto the cell monolayers and scrape with a rubber policeman. Skip the downstream centrifugation step and proceed with the homogenization.

5. If RNA quality controls reveal a poor enrichment of mRNA over rRNA, we recommend (a) increasing the volume of lysis buffer in **step 5**, (b) reducing the amount of beads, or (c) performing the hybridization at room temperature. These alterations of the original protocol may help to reduce nonspecific adherence to the oligo (dT) beads.
6. During homogenization maintain a good flow rate of the lysate through the narrow needle applying constant pressure on the syringe plunger. If the sample is still very viscous in spite of the three rounds of homogenization, increase either the number of homogenization cycles or the lysis volume.
7. At this step, it is possible to freeze the sample at –70–80 °C, although we recommend to avoid unnecessary freezing and, if possible, to proceed with the oligo (dT) capture after homogenization.
8. If RNA interactome capture is performed successfully, a halo will be visible around the bead pellet while washing with buffers 1, 2, and 3 in a UV cross-linking-dependent manner. Early appearance of the halo correlates with high protein content in the oligo (dT) pull down.
9. If the purification leads to a significant loss of magnetic beads, we recommend to add 0.025 % NP-40 (Igepal) to the buffers 1 and 2. Note that addition of detergent in these buffers will prevent the generation of halo (*see* **Note 8**). Avoid the use of NP-40 in buffer 3 since this detergent will impair downstream mass spectrometric analyses.
10. Removal of residual beads is key to prevent the blockage of the filter unit in downstream steps.
11. Protein can be concentrated by alternative methods such as ethanol or TCA precipitation. Nevertheless, we recommend to test these protocols for potential effects on downstream mass spectrometric analyses.

Acknowledgements

A.C. acknowledges support by the MRC Career Development Award Ref: MR/L019434/1. M.W.H. acknowledges support by the ERC Advanced Grant ERC-2011-ADG_20110310 and by a grant from the Virtual Liver Network of the German Ministry for Science and Education. A.C. is funded by the MRC Career Development Award number MR/L019434/1

References

1. Scherrer T, Mittal N, Janga SC, Gerber AP (2010) A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS One 5, e15499
2. Tsvetanova NG, Klass DM, Salzman J, Brown PO (2010) Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. PLoS One 5:e12671, 12610.11371/journal.pone.0012671
3. Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 30:1427–1464
4. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406
5. Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690
6. Kwon SC, Yi H, Eichelbaum K, Fohr S, Fischer B, You KT, Castello A, Krijgsveld J, Hentze MW, Kim VN (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol 20:1122–1130
7. Mitchell SF, Jain S, She M, Parker R (2013) Global analysis of yeast mRNPs. Nat Struct Mol Biol 20:127–133
8. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
9. Pashev IG, Dimitrov SI, Angelov D (1991) Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. Trends Biochem Sci 16:323–326
10. Suchanek M, Radzikowska A, Thiele C (2005) Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat Methods 2:261–267
11. Castello A, Horos R, Strein C, Fischer B, Eichelbaum K, Steinmetz LM, Krijgsveld J, Hentze MW (2013) System-wide identification of RNA-binding proteins by interactome capture. Nat Protoc 8:491–500
12. Strein C, Alleaume AM, Rothbauer U, Hentze MW, Castello A (2014) A versatile assay for RNA-binding proteins in living cells. RNA 20:721–731

Chapter 9

Identifying RBP Targets with RIP-seq

Hans-Herman Wessels, Antje Hirsekorn, Uwe Ohler, and Neelanjan Mukherjee

Abstract

Throughout their lifetime RNA molecules interact with a variety of RNA-binding proteins (RBPs). RBPs control gene expression by regulating splicing, polyadenylation, editing, transport, stability, and translation of RNA. There are ~1500 RBPs encoded by the human genome and recent studies have detected ~1100 proteins directly interacting with polyadenylated RNA. Identifying the RNAs bound by RBPs will continue to provide important insights into the regulation of gene expression.

Key words Ribonucleoprotein, Immunoprecipitation, RNA-binding protein, RNA-seq

1 Introduction

One of the challenges to assess the impact of RBPs on posttranscriptional gene regulation is identifying and quantifying the RNA and protein components of ribonucleoprotein complexes (RNPs) in a cellular context. The human genome is estimated to contain 1542 RBPs [1]. Recent studies indicated that approximately 800 proteins are bound to polyadenylated RNA within a single human cell type [2, 3]. Early approaches to identify RBP targets *en masse* independent of physiological context involved in vitro selection against limited RNA libraries [4]. The onset of microarray technology allowed for assessing RBP binding to RNA in a genome-wide manner [5, 6]. In these assays, RNP complexes are immunoprecipitated (RIP) from cell lysates. Associated RNAs are then isolated from these RNP complexes and interrogated with either microarray (RIP-chip) or sequencing (RIP-seq) technology. Unlike UV cross-linking and immunoprecipitation (CLIP-seq) methods [7], RIP-seq allows for the detection of RNA components of RNPs that are not directly bound to the RBP of interest. This is particularly relevant when interrogating multicomponent RNPs such as the exon junction complex [8]. RIP-seq provides

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_9,

whole transcript-level binding information rather than the site-level resolution of CLIP-seq methods. Comparisons have shown that RNAs with more CLIP-defined binding sites are more likely to be enriched in the RIP, suggesting that RIP identifies the more stably associated RNAs [9].

Here we describe the procedure of RNP immunoprecipitation followed by high-throughput sequencing (RIP-SEQ). This protocol is based on the RIP-chip protocol published by Keene and colleagues [5], but with a few modifications, additional steps describing quality control, and adaptations for RNA-seq. We lead you through sample preparation, in which the tissue or cell culture sample is harvested while preserving RNP:RNA interactions, coating beads with antibody, and RNP immunoprecipitation. Subsequently, RNA is extracted and utilized for qRT-PCR and library preparation and sequencing. The RIP-seq protocol is relatively short, simple, and does not require much specialized laboratory equipment. Its success relies on proper optimization of immunoprecipitation conditions. The main goal of these optimization experiments is to increase the significance and validity of obtained results by balancing high immunoprecipitation stringency and maximization of the signal-to-noise ratio of specifically bound RNA over background RNA. While post-lysis reassociation has been reported using a variant of this protocol for individual RNAs [10], we, and many others, have found no evidence for this using this protocol with the numerous RBPs and ELAVL1/HuR in particular [9]. Notes throughout the protocol give ideas for optimization of each step. Important controls include western blot and real-time PCR experiments described in Subheading 3.4. In this protocol we isolate and quantify RNAs associated to human ELAVL1 protein in HEK293 cells.

2 Materials

RNA samples must be handled cautiously to protect them from degradation caused by nucleases and heat. Always wear gloves. Keep samples at 4 °C as much as possible. Clean the workbench with nuclease-inhibitor solutions like RNase AWAY or equivalent. All solutions and buffers used for handling RNA samples should be prepared with pure and nuclease-free water, and processed with nuclease-free filter tips in nuclease-free low-retention reaction tubes.

2.1 Tissue Cell Culture Components

1. Human embryonic kidney cell line (HEK293) or Flp-In™ T-REx™ 293 Cells for the generation of cells with inducible expression epitope-tagged proteins. Here we are using stable FlpIn-HEK293-TO/FLAG/HA-ELAVL1 (*see* **Note 1**).

2. Cell growth medium: DMEM, 10 % fetal bovine serum (FBS), and 1 % penicillin/streptomycin for maintaining parental HEK293 cells.
3. Hygromycin: Add 100 μg/ml to the growth medium of FlpIn-HEK293-TO/FLAG/HA cells in order to maintain the stably integrated gene of interest.
4. Doxycycline: The expression of the epitope-tagged ELAVL1 protein is induced by adding doxycycline to the growth medium (final concentration 1 μg/ml; 1:1.000 v/v of 1 mg/ml doxycycline stock solution (*see* **Note 2**)).

2.2 Sample Collection Components

1. Phosphate-buffered saline (PBS) for cell culture (pH 7.4): 10× PBS contains potassium phosphate monobasic (KH_2PO_4) 1.440 g/L, sodium chloride (NaCl) 90 g/L, sodium phosphate dibasic (Na_2HPO_4-$7H_2O$) 7.950 g/L.
2. Polysome lysis buffer (PLB) (*see* **Note 3**): 100 mM Potassium chloride (KCl), 5 mM magnesium chloride ($MgCl_2$), 10 mM HEPES (pH 7.0), 0.5 % NP40, 1 mM dithiothreitol (DTT), 100 U/ml RNase Out, 1× complete proteinase inhibitor cocktail.

To prepare 5 ml PLB add 50 μL of 1 M HEPES (pH 7.0), 250 μL of 2 M KCl, 25 μL of 1 M $MgCl_2$, and 25 μL of Nonidet P- 40 (NP40) to 4.3775 ml of nuclease-free H_2O. Add 50 μL 1 M DTT, 12.5 μL RNase Out, and 200 μL of protease inhibitor cocktail (dissolved tablets according to the manufacturer's instructions) at the time of use.

2.3 RNP Immunoprecipitation Components

1. Protein A/G dynabeads: This choice depends on the species and IgG type of the antibody to be conjugated (*see* https://www.neb.com/tools-and-resources/selection-charts/affinity-of-protein-ag-for-igg-types-from-different-species).
2. Antibodies: Preferably monoclonal antibody recognizing RNA-binding protein (RBP) of interest, and an isotype-matched control antibody (here: mouse anti-FLAG M2 (Sigma F1804-200UG), and mouse anti-IgG1 (Thermo Scientific MA110406)).
3. NT2 buffer (*see* **Note 4**): 50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1 mM $MgCl_2$, 0.05 % NP40.

   To make 50 ml of NT2 buffer, add 2.5 ml of 1 M Tris (pH 7.4), 1.5 ml of 5 M NaCl, 25 μL of 2 M $MgCl_2$, and 25 μL of NP40 to 45.95 ml nuclease-free H_2O.
4. RNAsin.
5. Dithiothreitol (DTT).
6. Ethylenediamine tetraacetic acid (EDTA).
7. Proteinase K.

8. TRIZOL.
9. Chloroform, isopropanol, and glycogen.
10. Or Direct-Zol RNA MiniPrep kit (ZYMO) (*see* **Note 18**).

3 Methods

3.1 Tissue Cell Culture and mRNP Lysate Collection

1. Maintain FlpIn-HEK293-TO/FLAG/HA-ELAVL1 cells in DMEM cell growth medium containing 100 μg/ml hygromycin at 37 °C and 5 % CO_2. Per sample two 15 cm dishes will be needed (**Note 5**).
2. Induce the expression of the ELAVL1 gene by adding doxycycline to a final concentration of 1 μg/ml overnight.
3. Decant the growth medium, place the dish on ice, wash the adherent cells with ice-cold PBS at least once, add ice-cold PBS, scrape the cells off the dish using a cell lifter, and transfer the cells in PBS into a pre-chilled 50 ml conical tube. Spin down the cells at 200 × *g* at 4 °C for 5 min. Wash the cell pellet with cold PBS, spin down again, and aspirate the supernatant as much as possible, without disturbing the cell pellet.
4. Weigh the wet cell pellet and resuspend it thoroughly in an approximately equal volume PLB buffer (1:1; v/w) by pipetting up and down. Allow the lysate to chill on ice for 5 min before immediate freezing and storing the pellet at −80 °C until you proceed the experiment (*see* **Notes 5** and **6**).

3.2 Antibody Coating of Protein A/G Beads

For the immunoprecipitation of FLAG/HA epitope-tagged ELAVL1 protein, we use magnetic protein G dynabeads and coat them manually with mouse anti-FLAG M2 antibody according to the manufacturer's descriptions.

1. Vortex the dynabeads for 30 s and transfer the 20–30 μl of beads per sample into a nuclease-free low-retention 1.5 ml tube (*see* **Note 7**). An experiment of three biological replicates requires 120–180 μl of bead slurry. Half will be devoted for the RBP of interest and the other half for the control RIP (*see* Subheading 3.3, **step 3** and **Note 8**).
2. Wash the beads three times with PLB. Collect the beads using a magnetic stand and resuspend beads by pipetting up and down with low-affinity filter tips.
3. Dissolve the antibody in twice the original bead volume PLB, mix by pipetting up and down, and add it to the beads (*see* **Note 9**).
4. Incubate the beads at 4 °C on a rotating wheel at 20 rpm overnight the day before the immunoprecipitation. (Alternatively, coated beads can be stored at 4 °C by adding 0.02 % sodium azide for up to a month.)

5. Immediately before usage, wash the antibody-coated beads five times with ice-cold NT2 buffer to remove unbound antibody and potential contaminants like RNases.
6. After the final wash, resuspend the designated amount of antibody-coated beads per sample in 850 µL ice-cold NT2 buffer. Add to each sample/tube 40 U of RNase inhibitors (1 µL RNAsin), 10 µL of 100 mM DTT, and EDTA to a final concentration of 20 mM and keep them on ice until the lysate samples are ready to proceed with the IP.

3.3 Immunoprecipitation Reaction and RNA Precipitation

1. Allow the cell lysate (*see* Subheading 3.1, **step 4**) to thaw on ice.
2. Take a 10 µL aliquot of each sample, add 90 µL of NT2 buffer, and store this aliquot on ice for immunoprecipitation western blot (IP WB) control. This sample serves to control for RBP loss by spin clearing the lysate in the next step. These aliquots will be used in Subheading 3.4, **step 1**.
3. Clear the lysate by centrifuging at 4 °C at 20,000 × *g* for 15 min to remove larger particles and cell debris.
4. Transfer the cleared lysate into a new pre-chilled microfuge tube and store it on ice. Be careful not to disturb the formed pellet. At least 220 µl of lysate will be collected.
5. Take two 10 µL aliquots of the cleared lysate of each sample and store them on ice. The first one represents total cellular RNA as input for the RIP and will be needed for a qRT-PCR control and subsequent RNA-seq library preparation. The total RNA input is crucial to quantify the enrichment of specifically RBP-bound transcripts. The second aliquot is used for IP WB control to determine the input RBP amount. Add 90 µL of NT2 buffer to both aliquots. This matches the total RNA and protein concentration to the subsequent immunoprecipitation steps.
6. If necessary, pre-clearing of lysate with beads may be used to reduce background (*see* **Note 10**).
7. For each sample add 100 µL of the cleared lysate to the beads coated with either the antibody recognizing the RBP and the control antibody from **step 6,** Subheading 3.3.
8. Slowly rotate the IP sample at 4 °C tumbling end over for 4 h (*see* **Note 11**).
9. Collect the beads using the magnetic stand and transfer the supernatant into a new tube (*see* **Note 12**). Take a 30 µL aliquot of the supernatant and store it on ice for later WB IP control. This aliquot serves as a control for estimating the IP efficiency.
10. Wash the beads five times with 1 ml of ice cold NT2 buffer by pipetting up and down with low-affinity filter tips (*see* **Notes 13–15**).

11. When you are at the last washing step, dissolve the beads in 1 ml of the washing buffer and take 20 μL of the beads slurry and store it on ice for IP WB control. This sample, together with the input and supernatant, serves to control for effective RBP immunoprecipitation and potential loss of RBP during the washing steps.
12. Resuspend the beads in 100 μL of NT2 buffer. Add 30 μg of proteinase K to release the RNP complexes from the beads. Incubate at 55 °C for 30 min, flicking the tube occasionally with the finger from time to time.
13. Release the RNA from the RNP complexes by adding 1 ml of TRIZOL reagent directly to the tube.
14. Isolate RNA following the manufacturer's descriptions from either TRIZOL (*see* **Notes 16** and **17**) or by using the column-based Direct-Zol RNA MiniPrep kit (*see* **Note 18**).
15. Resuspend (*see* **Note 16**) or elute the RNA in maximally 30 μL RNase-free water (*see* **Note 19**) for subsequent RNA quantification (*see* **Note 20**). RNA can be stored at 80 °C for months.

3.4 RNP Immunoprecipitation Controls

Before proceeding to RNA-seq library preparation it is important to perform control experiments to assess enrichment of both the RBP (western blot) and RNA components (qRT-PCR) of the RNP being immunoprecipitated. The qRT-PCR experiments require knowledge of at least one RNA known to be a component of the RNP complex of interest and at least one RNA known not to be a component of that RNP complex.

1. Run the collected lysate aliquots (cleared lysate/input, supernatant, (optional: IP wash step) and IP/bead slurry) on an SDS-PAGE followed by western blotting. Assay the membrane for your protein of interest (*see* **Note 21** and Fig. 1).
2. If you have information about candidate transcripts bound to your RBP of interest, perform qRT-PCR on your RNA samples (total RNA input, IP, and isotype-matched IP control) (*see* **Note 22** and Fig. 2a, b).

3.5 Library Preparation

1. All RNA samples can directly be applied to standard long RNA-seq protocols. Here we utilize the NEXTflex™ Rapid Directional qRNA-Seq Kit (BIOO), which requires 10–100 ng of RNA as input. Using this protocol in the RIP samples we typically collect 0.5–1.5 μg of RNA. Thus, even if the initial material for the lysate is substantially less than described in this protocol it is still possible to generate high quality libraries (Fig. 3).
2. If the RBP is not expected to bind rRNA, it is advisable to deplete ribosomal RNA (rRNA) from all RNA samples using Ribo-Zero (epicentre) previous to generating libraries.

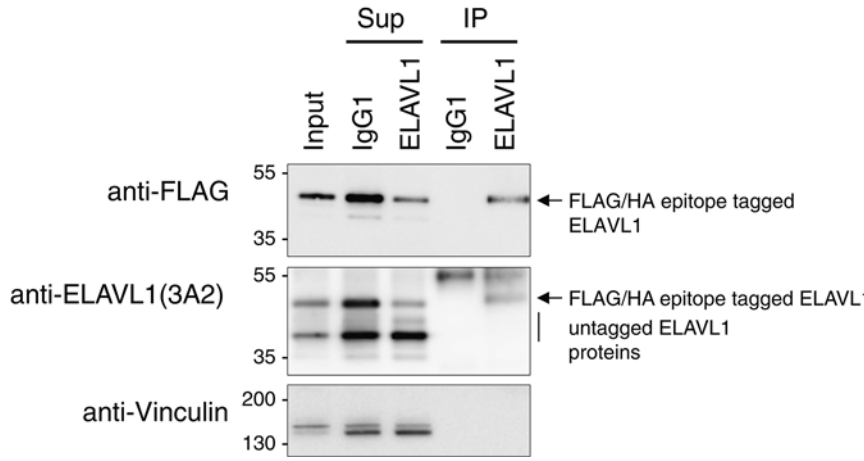

Fig. 1 Optimizing RIP: western blot controls. Western blot of input, supernatant, and IP samples probed with anti-FLAG, anti-ELAVL1, and anti-vinculin (loading control). Note the absence of FLAG/HA-ELAVL1 in the IgG IP and the depletion of FLAG/HA-ELAVL1 in the FLAG IP supernatant. Input: 10 μL of input lysate was diluted 10× in PLB to make it comparable to the Sup (=10 %). Of this 5 % was loaded (0.5 % of total input). Sup: 50 μl of 10× diluted input (=5 %). Of this 10 % was loaded (0.5 % of total input). IP: 10 μl of beads diluted in 1 mL IP reaction mix was taken (=1 %). Of this 5 μl was loaded (0.5 % of total input)

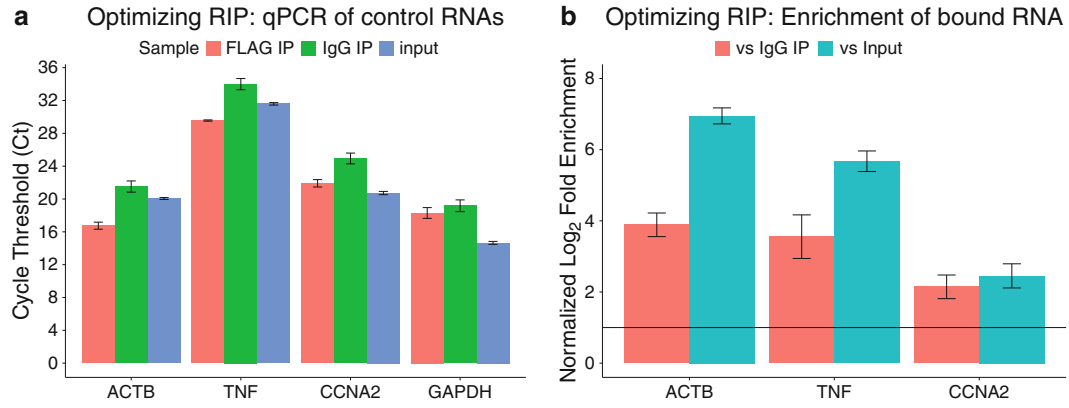

Fig. 2 Optimizing RIP: qRT-PCR controls. (**a**) Cycle threshold (Ct) values in FLAG IP, IgG IP, and input for three known target RNAs (ACTB, TNF, and CCNA2) and one nontarget RNA (GAPDH). For the known targets, notice the lower Ct values indicating higher RNA levels, in the FLAG IP compared to the IgG IP as well as the input. For the nontarget, the RNA is detected at a later cycle number in the input than either IP. Furthermore, for GAPDH the Ct values are comparable in both IP samples indicating this is a suitable normalization control. (**b**) All Ct values were normalized using GAPDH, and then FLAG IP samples were compared to IgG and input. The y-axis is log scaled, so the fold enrichment of the targets ranges from 4× to > 100×

3.6 Computational Analysis

1. The bioinformatics analysis of this data is extremely important. However, it is not the focus of this protocol. Numerous software are available for analyzing RIP-seq data including ASPeak, RIPseeker, and Piranha [11–13].

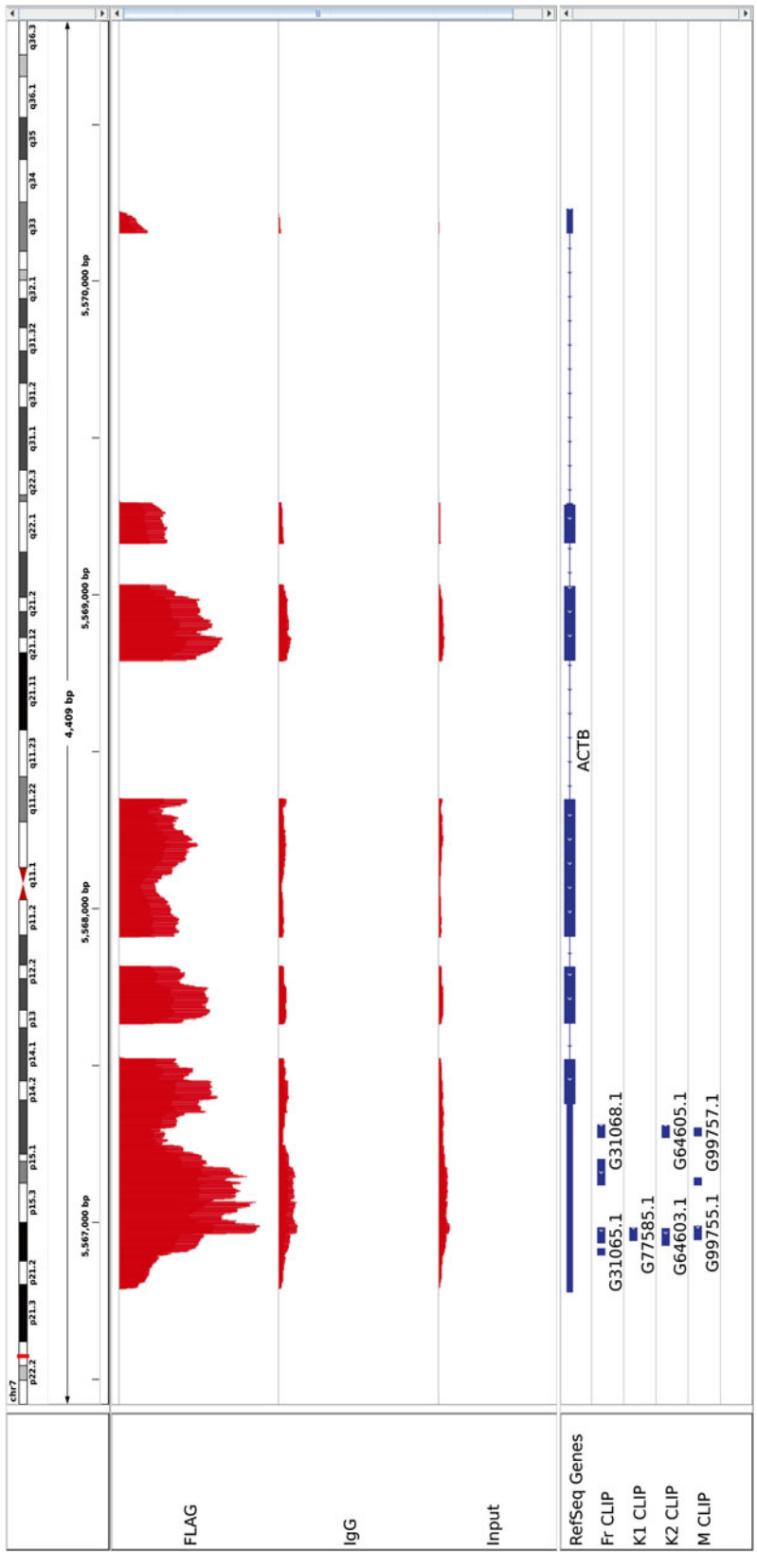

Fig. 3 Expected outcome: Enrichment of ACTB in FLAG IP relative to IgG IP and input. IGV browser snapshot of RIP-seq data for ACTB depicts normalized RIP-seq coverage (*red*) for the FLAG IP (*top*), IgG IP (*middle*), and input RNA (*bottom*). PAR-CLIP-defined binding sites (*blue*) of ELAVL/HuR from 4 independent libraries all performed in HEK293 cells

4 Notes

1. RIP experiments can be performed in tissue samples from living organisms for assaying the RBP-binding partners in their physiological context. However, it is recommended to study your RBP of interest first in a cell line in order to optimize experiment conditions, particularly if an IP-grade antibody does not exist and the RBP must be epitope tagged.
2. If using an inducible expression system, it is important to control the expression level of your protein of interest. Keep expression levels similar to physiological expression.
3. PLB allows for very mild cell lyses, which does not aim for complete organelle lysis. For complete lysis the PLB will need to be adjusted in its composition by increasing detergent (NP40 or Triton-X) concentration, or by physical disruption. Take caution using denaturing reagents like SDS, since native conditions are required for the RIP. Physical disruption may become important for tissue samples, or samples with cell walls. Note that sonication can disrupt endogenous protein:RNA interactions.
4. NT2 buffer allows for mild washing of the immunoprecipitated RNPs. These washing conditions need to be optimized carefully in order to wash as stringent as possible to remove background, but as mild as necessary retain the RNP of interest. These adjustments can involve higher salt concentrations, higher concentrations of NP40, or the addition of urea, SDS, or DTT.
5. It is advisable to have at least three biological replicates for each RIP experiment. Each replicate/sample requires a minimum of 250 mg wet cell pellet, which results in 500 μl total lysate volume for a complete sample. Clearing the lysate leads to an approximate loss of 50 % of the initial lysate volume. After clearing, 220 μl lysate should remain (10 μl for total RNA input, 10 μl for input for IP WB, and 2× 100 μL for the RNP IP and control IP, respectively). For the HEK293 cells used here, 250 mg wet cell pellet correspond to approximately 2×15 cm^2 dishes 80 % confluent. In most cases this will yield much more RNA than necessary for library preparation; however this will strongly depend on the RNP being studied.
6. The immediate freezing is necessary for complete lysis and to avoid post-lysis reassortment of target RNAs and nonspecific binding. Additional freeze-thaw cycles should be avoided to prevent protein and RNA degradation. It is advisable to store your sample lysates in either 500 μl aliquots, or 220–250 μl aliquots of cleared lysate.

7. It is recommended to choose a small volume of beads per sample, in order to reduce chances of nonspecific background binding, but sufficiently high for proper processing of the samples, normally 20–30 μL.
8. In parallel, an isotype-matched control antibody should be used to assess background RNA binding.
9. The amount of antibody used should be carefully determined beforehand. An appropriate amount would be the least amount necessary to deplete the majority of the available protein. If the antibody titer is too high, it may increase nonspecific binding and background RNA signal. If it is too low, the IP RNA yield might be insufficient.
10. RNA molecules can nonspecifically bind to the bead surface making it difficult to distinguish truly bound transcripts from background. One way to overcome this issue is to pre-clear the lysate. Incubate the lysate on the rotating wheel with uncoated beads, similar to the immunoprecipitation. This will remove RNA species, which nonspecifically bind to the beads, before the actual IP is done. However, pre-clearing may reduce signal.
11. The immunoprecipitation step can be performed at 4 °C over night or at room temperature conditions (18 °C–25 °C) for a shorter time period (15–60 min). This has to be optimized for the specific antibodies and RBPs used.
12. The supernatant can be stored at –80 °C and be used to IP other RBPs from the same sample. This can be useful to explore competition of RBPs for an overlapping pool of target RNAs.
13. During IP optimizations, it is recommended to collect a 30 μL aliquot of the washing steps for the IP WB. Detection of your RBP in these samples indicates loss of bound RBP due to excessively stringent washing conditions.
14. It is critical to stringently wash the immunoprecipitated RNPs. These washing conditions need to be optimized carefully in order to wash as stringent as possible to remove background, but as mild as necessary retain the RNP of interest. These adjustments can involve higher concentration of salts or NP40 detergent, and/or the addition of urea, SDS, or DTT.
15. All steps, including washing the beads, must be carried out on ice as much as possible. In addition, work quickly and do not to let the beads dry in between the individual wash steps. Higher number of samples might require processing your samples in batches, or one by one.
16. The standard TRIZOL procedure includes an RNA precipitation step. Since the IP RNA yield is relatively low, the precipitated RNA pellet may be difficult to see. Adding glycogen as a carrier to the precipitation aids to recover RNA and helps to make the recovered RNA pellet visible in the tube.

17. TRIZOL is phenol based and requires the use of chloroform. RNA isolation should be done in a fuming hood.
18. If using a column-based RNA isolation method, it is important to know the RNA fragment size cutoff of the columns used. The kit used here isolates RNA down to 17 nucleotides in length and is therefore appropriate for microRNA and other small RNA species.
19. The resuspension/elution volume can be adjusted to the expected yield. This will effect RNA concentrations for subsequent RNA sample handling.
20. For accurate RNA quantification of the RIP samples use sensitive methods (e.g., Qubit RNA HS Assay for the Qubit 2.0 Fluorometer, Life). Standard RNA quantification is suitable for the input RNA.
21. This western blot control helps determine if the RBP of interest was sufficiently immunoprecipitated from the sample. Expect to see a depleted signal in your supernatant and enrichment in your IP/bead fraction. Complete RBP depletion in the supernatant fraction is not required (*see* **Note 8**). For the nonspecific isotype control IP, there should not be a signal in the IP/bead fraction for your RBP of interest. While you are optimizing IP conditions, it is recommended to include an aliquot of the IP wash step to assess RBP loss due to too harsh buffer conditions.
22. Quantify the RNA samples and reverse transcribe approximately 1–10 ng of each RNA sample into cDNA using iScript (Biorad). Assay enrichment of candidate-binding transcript and several transcripts of potentially non-bound transcripts using qRT-PCR. It is critical to observe enrichment of the candidate-bound transcript in the RNP IP sample over input and isotype-matched control IP sample.

References

1. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
2. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406
3. Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690
4. Gao FB, Carson CC, Levine T, Keene JD (1994) Selection of a subset of mRNAs from combinatorial 3' untranslated region libraries using neuronal RNA-binding protein Hel-N1. Proc Natl Acad Sci U S A 91:11207–11211
5. Keene JD, Komisarow JM, Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc 1:302–307
6. Tenenbaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in

messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci U S A 97:14085–14090

7. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215
8. Singh G, Ricci EP, Moore MJ (2014) RIPiT-Seq: a high-throughput approach for footprinting RNA:protein complexes. Methods 65:320–332
9. Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M Jr, Tuschl T, Ohler U, Keene JD (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol Cell 43:327–339
10. Mili S, Steitz JA (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA 10:1692–1694
11. Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD (2012) Site identification in high-throughput RNA-protein interaction data. Bioinformatics (Oxford, England) 28:3013–3020
12. Kucukural A, Ozadam H, Singh G, Moore MJ, Cenik C (2013) ASPeak: an abundance sensitive peak detection algorithm for RIP-Seq. Bioinformatics (Oxford, England) 29:2485–2486
13. Li Y, Zhao DY, Greenblatt JF, Zhang Z (2013) RIPSeeker: a statistical package for identifying protein-associated transcripts from RIP-seq experiments. Nucleic Acids Res 41, e94

Chapter 10

PAR-CLIP: A Method for Transcriptome-Wide Identification of RNA Binding Protein Interaction Sites

Charles Danan, Sudhir Manickavel, and Markus Hafner

Abstract

During *p*ost-*t*ranscriptional *g*ene *r*egulation (PTGR), *R*NA *b*inding *p*roteins (RBPs) interact with all classes of RNA to control RNA maturation, stability, transport, and translation. Here, we describe *P*hoto*a*ctivatable-*R*ibonucleoside-Enhanced *C*ross*l*inking and *I*mmuno*p*recipitation (PAR-CLIP), a transcriptome-scale method for identifying RBP binding sites on target RNAs with nucleotide-level resolution. This method is readily applicable to any protein directly contacting RNA, including RBPs that are predicted to bind in a sequence- or structure-dependent manner at discrete *R*NA *r*ecognition *e*lements (RREs), and those that are thought to bind transiently, such as RNA polymerases or helicases.

Key words RNA-binding protein (RBP), RNA, Photoactivatable ribonucleoside enhanced cross-linking and immunoprecipitation (PAR-CLIP), Cross-linking and immunoprecipitation (CLIP), Posttranscriptional gene regulation (PTGR), RNA recognition element (RRE), Noncoding RNA, mRNA, Binding site

1 Introduction

All classes of RNA are subject to *post*transcriptional *g*ene *r*egulation (PTGR), including splicing, 5′- and 3′-end-modification, editing, transport, translation, and degradation [1–3]. These processes are critical for the regulation of protein-coding *m*essenger *R*NA (mRNA), as well as for the biogenesis and function of *n*on-*c*oding *R*NAs (ncRNAs, e.g., ribosomal RNA, microRNA, small interfering RNA, etc.), which themselves have a wide range of gene-regulatory functions [4]. PTGR is coordinated by the actions of *ri*bo*nucleop*roteins (RNPs), protein–RNA complexes composed of one or more *R*NA *b*inding *p*roteins (RBPs), and associated coding or noncoding RNAs.

The fundamental importance of PTGR is reflected in analyses of abundance, expression patterns, and evolutionary conservation

* These authors contributed equally to this work and appear in alphabetical order.

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_10, © Springer Science+Business Media New York 2016

of RBPs. In human cell lines and tissues, approximately 20 % of the protein-coding transcriptome is comprised of RBPs, making RBPs more abundant than most other classes of proteins. The low tissue-specificity and deep evolutionary conservation of most RBP families suggests that many PTGR processes are ancient and equally essential for all cells [4]. Dysregulation of PTGR is observed in a wide variety of human pathologies, ranging from musculoskeletal and autoimmune disorders, to neurodegenerative disease, to essentially all forms of cancer [5–7].

Dissection of PTGR networks requires the careful characterization of the molecular interactions of RBPs with their RNA ligands and other binding partners, but this effort is complicated by the vast size of PTGR networks. In humans, there are approximately 1500 proteins containing identified *R*NA *b*inding *d*omains (RBDs), and over 20,000 protein-coding mRNAs in addition to the thousands of diverse noncoding RNAs [8]. Each RBP binds at defined sequence and structural elements termed *R*NA *r*ecognition *e*lements (RREs). However, RREs are short and partially degenerate, confounding reliable computational predictions and sparking the need for experimental methods to comprehensively identify RREs on a transcriptome-wide scale [9].

Traditionally, RREs were characterized individually in a reductive process; sequences from known RNA targets were analyzed and then putative RREs were biochemically validated. Characterization of RNPs on a transcriptome-wide scale first became possible using RNP Immunoprecipitation (RIP) followed by comprehensive identification and quantification of recovered RNAs by microarray or next generation sequencing analysis (RIP-Chip or RIP-seq) [10]. However, RIP methods are limited to the analysis of kinetically stable interactions. Furthermore, the RRE needs to be inferred computationally from the sequence of the long recovered RNAs, which is only successful for RREs with high information content [11, 12].

The recently introduced *C*ross*L*inking and *I*mmuno*p*recipitation (CLIP) approaches use UV light to covalently cross-link RBPs with their RNA targets at the site of interaction. The covalent bond between the RBP and target RNAs allows for limited RNase digestions to trim the RNA to the footprint protected by the RBP, as well as additional stringent purification steps after IP, including denaturing polyacrylamide gel-electrophoresis and blotting onto nitrocellulose membranes. The recovered RNA segments can then be sequenced using next-generation sequencing technologies to reveal target transcripts and RREs on a transcriptome-wide scale [13, 14].

Here we provide a step-by-step protocol for *Photoactivatable-Ribonucleoside-Enhanced* CLIP (PAR-CLIP) (Fig. 1). In PAR-CLIP, photoactivatable ribonucleosides—4-thiouridine (4SU), or more rarely, 6-thioguanine (6SG)—are incorporated into nascent RNA transcripts. The labeled RNAs are excited in living cells with UVA or UVB light (>310 nm) and yield photoadducts with

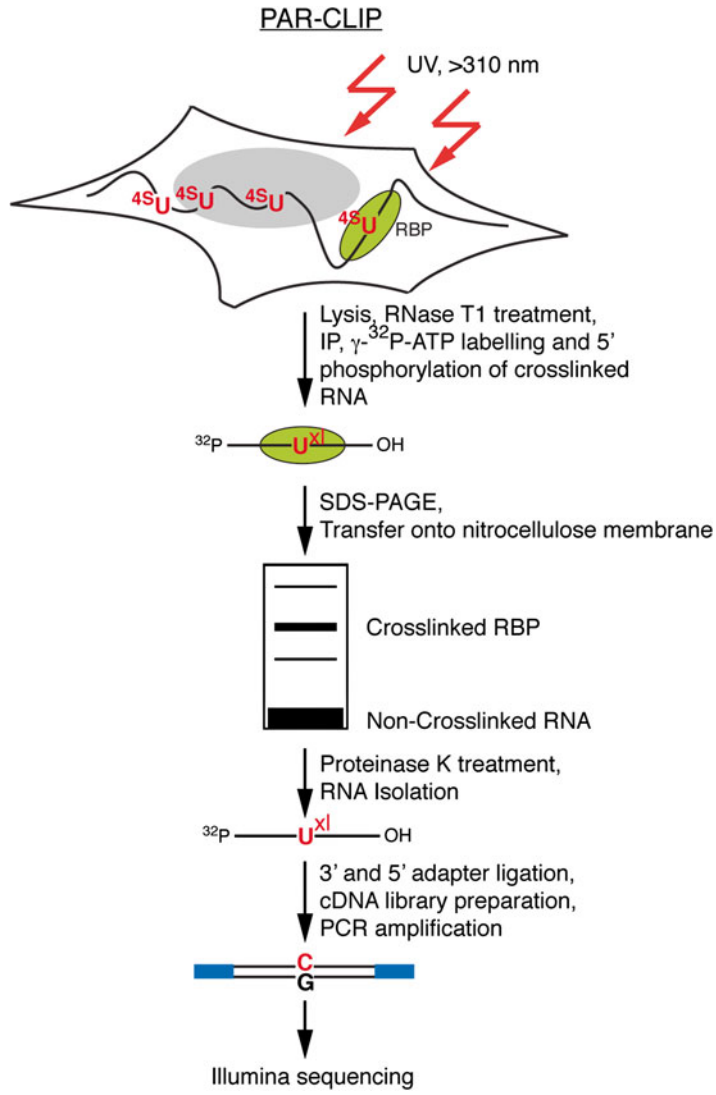

Fig. 1 Outline of the PAR-CLIP methodology. PAR-CLIP begins with incorporation of photoactivatable thioribonucleosides into nascent transcripts followed by cross-linking with long-wavelength >310 nm UV. Cross-linked RNA–RBP complexes are isolated by immunoprecipitation and further purified by SDS-PAGE. After recovery from the purified radioactive band, the RNA is carried through a small RNA cDNA library preparation protocol for sequencing. Reverse transcription of cross-linked RNA with incorporated photoactivatable thioribonucleosides, followed by PCR amplification, leads to a characteristic mutation (T-to-C when using 4SU and G-to-A when using 6SG) that is used to identify the RNA recognition elements

interacting RBPs. Besides an increased cross-linking efficiency compared to 254 nm CLIP, a key feature of PAR-CLIP is a characteristic mutation (T-to-C for 4SU and G-to-A for 6SG) introduced during reverse transcription at the position of cross-linking. This

mutation pinpoints the sites of RNA–RBP interaction with nucleotide resolution. And, more importantly, it enables the user to computationally remove the ubiquitous background of co-purifying fragments of cellular RNAs that otherwise may be misinterpreted as signal [15]. The resulting detailed interaction maps will further our understanding of the mechanisms underlying the pathologic dysregulation of PTGR components. This information can also be integrated with emerging data from other large-scale sequencing efforts to interrogate whether variations in binding sites contribute to phenotypic variations or complex genetic disease.

The following guide covers all experimental steps of PAR-CLIP and cDNA library construction and touches on a number of aspects of the data analysis.

2 Materials

1. 4-Thiouridine (4SU) stock solution (1 M): 260.27 mg 4SU in 1 ml DMSO.
2. 1× NP40 lysis buffer: 50 mM HEPES, pH 7.5, 150 mM KCl, 2 mM EDTA, 1 mM NaF, 0.5 % (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitory cocktail (Roche).
3. High-salt wash buffer: 50 mM HEPES–KOH, pH 7.5, 500 mM KCl, 0.05 % (v/v) NP40, 0.5 mM DTT, complete EDTA-free protease inhibitor cocktail (Roche).
4. Dephosphorylation Buffer: 50 mM Tris–HCl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT.
5. Polynucleotide Kinase (PNK) Buffer without DTT: 50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$.
6. PNK Buffer with DTT: 50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT.
7. SDS PAGE Loading Buffer: 10 % glycerol (v/v), 50 MM Tris–HCl, pH 6.8, 2 mM EDTA, 2 % SDS (w/v), 100 mM DTT, 0.1 % bromophenol blue.
8. 1× Transfer Buffer with Methanol: 1× NuPAGE Transfer Buffer, 20 % MeOH.
9. 2× Proteinase K Buffer: 100 mM Tris–HCl, pH 7.5 150 mM NaCl, 12.5 mM EDTA, 2 % (w/v) SDS.
10. Acidic Phenol–Chloroform–IAA: 25 ml acidic phenol, 24 ml chloroform, 1 ml isoamyl alcohol, pH 4.2.
11. 10× RNA Ligase Buffer without ATP: 0.5 M Tris–HCl, pH 7.6, 0.1 M $MgCl_2$, 0.1 M 2-mercaptoethanol, 1 mg/ml acetylated BSA (Sigma, B-8894).
12. 10× RNA Ligase Buffer with ATP: 0.5 M Tris–HCl, pH 7.6, 0.1 M $MgCl_2$, 0.1 M 2-mercaptoethanol, 1 mg/ml acetylated BSA (Sigma, B-8894), 2 mM ATP.

13. Formamide Gel Loading Dye: 50 mM EDTA, 0.05 % (w/v) bromophenol blue, formamide ad 100 %.
14. 10× dNTP Solution: 2 mM dATP, 2 mM dCTP, 2 mM dGTP, 2 mM dTTP.
15. 10× PCR Buffer: 100 mM Tris–HCl, pH 8.0, 500 mM KCl, 1 % Triton-X-100, 20 mM $MgCl_2$, 10 mM 2-mercaptoethanol.
16. Dynabeads Protein G: Invitrogen, 100.03D/04D.
17. 15 ml Falcon Centrifuge Tubes: Fisher Scientific.
18. 1.5 ml DNA LoBind Tubes: Eppendorf.
19. RNase T1 (1000 U/μl): Fermentas, EN0541.
20. Calf Intestinal Alkaline Phosphatase (10,000 U/ml): New England Biolabs (NEB), M0290.
21. T4 Polynucleotide Kinase (10,000 U/ml): NEB, M0201.
22. γ-32P-ATP, 10 mCi/ml, 1.6 μM: Perkin Elmer, NEG002Z001MC.
23. NuPAGE Novex 4–12 % BT Midi 1.0 gel: Invitrogen.
24. 20× NuPAGE MOPS running buffer: Invitrogen.
25. Protein Size Marker: Bio-Rad, 161-0374.
26. 20× NuPAGE Transfer Buffer: Invitrogen.
27. 0.45 μm Nitrocellulose Membrane: Invitrogen.
28. Proteinase K (Powder): Roche, 03 115 879 001.
29. GlycoBlue, 10 mg/ml: Ambion.
30. Truncated and mutated RNA Ligase 2, T4 Rnl2 (1-249) K227Q, 1 mg/ml: NEB, M0351, plasmid for recombinant expression can also be obtained at addgene.org.
31. T4 RNA Ligase (10 U/μl): Thermo Scientific.
32. SuperScript III Reverse Transcriptase: Invitrogen, 18080-044.
33. Taq DNA Polymerase, 5 U/μl: Various Suppliers.
34. MinElute Gel Extraction Kit: Qiagen.
35. Pre-adenylated 3′ Adapter (DNA): AppTCGTATGCCGTC TTCTGCTTGT.
36. 5′ Adapter (RNA): GUUCAGAGUUCUACAGUCCGAC GAUC.
37. 3′ Primer: CAAGCAGAAGACGGCATACGA.
38. 5′ Primer: AATGATACGGCGACCACCGACAGGTTC AGAGTTCTACAGTCCGA.
39. RNA Size Marker, 19 nt: 5′ pCGUACGCGGUUUAAACGA.
40. RNA Size Marker, 35 nt: 5′ pCUCAUCUUGGU CGUACGUACGCGGAAUAGUUUAAACUGU.

3 Methods

Before beginning PAR-CLIP, please *see* **Notes 1–7** for essential preparatory steps.

3.1 Preparation of UV-Crosslinked RNPs

3.1.1 Expanding Cells

1. Expand cells in appropriate growth medium in 15-cm plates. As a starting point, we recommend using a number of cells that will result in 1.5–3 ml of wet cell pellet. For HEK293 cells, approximately $100–200 \times 10^6$ cells will result from 10 to 20 15-cm plates. Grow cells to approximately 80 % confluency.
2. 16 h before crosslinking, add 4SU to a final concentration of 100 μM (1:1000 v/v of a 1 M 4SU stock solution) directly to the cell culture medium (*see* **Note 8**).

3.1.2 UV-Crosslinking

For adherent cells

1. Aspirate or pour off media from plates (*see* **Note 9**).
2. Irradiate cells uncovered with a dose of 0.15 J/cm^2 of >310 nm UV light in a Spectrolinker XL-1500 (Spectronics Corporation) equipped with >310 nm light bulbs or similar device.
3. Cover cells in 1 ml PBS and scrape cells off with a rubber policeman. Transfer the cell suspension to 50 ml centrifugation tubes and collect by centrifugation at $500 \times g$ at 4 °C for 5 min. Discard the supernatant.

Stopping point: If you do not want to continue directly with cell lysis and immunoprecipitation, snap freeze the cell pellet in liquid nitrogen and store at −80 °C. Cell pellets can be stored for at least 12 months.

For cells grown in suspension

1. Collect cells by centrifugation at $500 \times g$ at 4 °C for 5 min. Aspirate or pour off media.
2. Take up cells in 10 ml PBS and transfer onto one 15-cm cell culture plate.
3. Irradiate uncovered with a dose of 0.2 J/cm^2 of >310 nm UV light in a Spectrolinker XL-1500 (Spectronics Corporation) equipped with >310 nm light bulbs or similar device.
4. Transfer cells into a 50 ml centrifugation tube and collect by centrifugation at $500 \times g$ for 5 min at 4 °C and discard the supernatant.

Stopping point: If you do not want to continue directly with cell lysis and immunoprecipitation, snap freeze the cell pellet in liquid nitrogen and store at −80 °C. Cell pellets can be stored for at least 12 months.

3.1.3 Cell Lysis and RNase T1 Digest

1. Take up cross-linked cell pellet in 3 volumes of 1× NP40 lysis buffer and incubate on ice for 10 min in a 15 ml centrifuge tube.
2. Clear cell lysate by centrifugation at 13,000×*g* at 4 °C for 15 min. In the meantime, begin to prepare the magnetic beads (*see* Section 3.2.1).
3. Transfer supernatant to a new 15 ml centrifuge tube. Discard the pellet.
4. Add RNase T1 to a final concentration of 1 U/μl and incubate at 22 °C for 15 min. Cool reaction subsequently for 5 min on ice before proceeding (*see* **Note 10**).
5. Keep a 100 μl aliquot of cell lysate and store at −20 °C to control for RBP expression in Subheading 3.2.4.

See **Note 11** *for guidelines on handling and washing of magnetic beads.*

3.2 Immunoprecipitation and Recovery of Crosslinked Target RNA Fragments

3.2.1 Preparation of Magnetic Beads

1. Transfer 20 μl of Protein G magnetic beads per ml of cell lysate to a 1.5 ml microcentrifuge tube (for a typical experiment, 120–200 μl of beads) (*see* **Note 12**).
2. Wash beads twice in 1 ml of PBS.
3. Resuspend beads in twice the volume of PBS relative to the original bead volume aliquotted.
4. Add 0.25 mg of antibody per ml original bead volume and incubate on a rotating wheel for 40 min at room temperature.
5. Wash beads twice in 1 ml of PBS to remove unbound antibody.
6. Resuspend beads in one original bead volume of PBS.

3.2.2 Immunoprecipitation (IP), Second RNase T1 Digestion, and Dephosphorylation

1. Add 20 μl of freshly prepared antibody-conjugated magnetic beads per 1 ml of partial RNase T1-treated cell lysate (from Subheading 3.1.3, **step 4**) and incubate in 15 ml centrifuge tubes on a rotating wheel for 1 h at 4 °C.
2. Collect magnetic beads on a magnetic particle collector for 15 ml centrifuge tubes.
3. Keep a 100 μl aliquot of supernatant and store at −20 °C to control for RBP depletion in Subheading 3.2.4. Discard the remaining supernatant.
4. Add 1 ml of 1× NP40 lysis buffer to the centrifugation tube and transfer the suspension to a 1.5 ml microcentrifuge tube (*see* **Note 13**).
5. Wash beads twice in 1 ml of 1× NP40 lysis buffer.
6. Take up cells in one original bead volume of 1× NP40 lysis buffer.

7. Add RNase T1 to a final concentration of 1 U/μl and incubate the bead suspension at 22 °C for 15 min. Cool subsequently on ice for 5 min.
8. Wash beads twice in 1 ml of 1× NP40 lysis buffer (*see* **Note 14**).
9. Wash beads twice in 400 μl of dephosphorylation buffer.
10. Resuspend beads in 1 original bead volume of dephosphorylation buffer.
11. Add calf intestinal alkaline phosphatase to a final concentration of 0.5 U/μl and incubate the suspension at 37 °C for 10 min.
12. Wash beads twice in 1 ml of 1× NP40 lysis buffer.
13. Wash beads twice in 1 ml of polynucleotide kinase (PNK) buffer without DTT (*see* **Note 15**).
14. Resuspend beads in 1 original bead volume of PNK buffer with DTT.

3.2.3 Radiolabeling of RNA Segments Crosslinked to Immunoprecipitated Proteins

1. To the bead suspension described above, add T4 polynucleotide kinase to 1 U/μl and γ-^{32}P-ATP to a final concentration of 0.5 μCi/μl (1.6 μM ATP) in one original bead volume. Incubate at 37 °C for 30 min.
2. Add nonradioactive ATP to obtain a final concentration of 100 μM and incubate at 37 °C for another 5 min.
3. Wash magnetic beads five times with 800 μl of PNK buffer without DTT. Store a 100 μL aliquot of radioactive wash waste for use as radioactive markers in future steps.
4. Resuspend the beads in 70 μl of SDS-PAGE loading buffer and incubate for 5 min in a heat block at 95 °C to denature and release the immunoprecipitated RNPs. Vortex and centrifuge briefly.
5. Remove the magnetic beads on the magnetic separator and transfer the supernatant (i.e., radiolabeled RNP immunoprecipitate) to a clean 1.5 ml microcentrifuge tube.

Stopping point: The sample can be stored at −20 °C for a prolonged period of time. However, the half-life of ^{32}P half life is 14.5 days, and we therefore recommend continuing with the protocol within 2 weeks.

3.2.4 SDS Polyacrylamide Gel Electrophoresis, Transfer, and Recovery of RNA from Nitrocellulose Membrane

1. Prepare a 4–12 % Bis-Tris polyacrylamide gel. We recommend using the first half of the gel for separation of the radiolabeled RNP IP and the second half for immunoblotting to control for RBP expression and IP efficiency. In the first half of the gel, load 40 μl of the radiolabeled RNP IP per well. Each RNP IP sample should be loaded adjacent to a ladder and there should be at least one lane distance between different samples. In the second half of the gel, load 10 μL of cell lysate (from Subheading 3.1.3,

step 5), 10 μL supernatant (from Subheading 3.2.2, **step 3**), and 2 μl of the radiolabeled IP.

2. Run the gel at 200 V for 40 min.
3. Using semidry blotting, transfer proteins onto a 0.45 μm nitrocellulose membrane in 1× transfer buffer at 2 mA/cm^2 current for 1 h.
4. Using a scalpel or razor blade, split the nitrocellulose membrane in two, separating the RNP IP samples from the samples for immunoblotting. Proceed to **step 5** with the RNP IP samples. With the lanes for testing IP and RBP expression, perform a Western blot to probe for your RBP or RBP-tag.
5. Label three corners of the membrane and each band of the protein length marker with 1 μl of radioactive wash waste from Subheading 3.2.3, **step 3**. Wrap the membrane in plastic film (e.g., Saran wrap) to avoid contamination of the phosphorimager screen.
6. Expose the membrane to a blanked phosphorimager screen for 1 h at room temperature and visualize on a phosphorimager. If the radioactivity of the recovered RNP is weak, you can expose the membrane for longer.

Stopping point: The membrane can be stored at −20 °C for a prolonged period of time.

7. Print the image from the phosphorimager onto an overhead projector transparency film; make sure the image is scaled to 100 % for printing. Align the transparency film printout on top of the membrane using the labeled corners for orientation.
8. Cutting through the transparency and the membrane directly beneath, excise the bands on the nitrocellulose membrane that correspond to the expected size of the RBP.
9. Cut the nitrocellulose excisions further by slicing them into ~5 smaller pieces. Transfer the pieces into a 1.5 ml low adhesion tube (e.g., siliconized or DNA LoBind tubes).

3.2.5 Proteinase K Digestion

1. Add 400 μl of 1× Proteinase K buffer to the nitrocellulose pieces followed by the addition of approximately 2 mg Proteinase K. Vortex, briefly centrifuge, and incubate at 55 °C for 1 h 30 min.
2. Extract the RNA by addition of 2 volumes of acidic phenol–chloroform–IAA (25:24:1, pH 4.0) directly to the Proteinase K digestion. Vortex for 15 s and centrifuge at >14,000 × *g* at 4 °C for 5 min. Remove the aqueous phase without disturbing the organic phase or interphase, and transfer the aqueous phase to a new 1.5 ml low adhesion microcentrifuge tube. If the organic or interphase is accidentally disturbed, centrifuge the sample again and reattempt.

3. Add 1 volume of chloroform to the recovered aqueous phase to remove residual phenol. Vortex for 15 s and centrifuge at >14,000 × *g* at 4 °C for 5 min. Remove the aqueous phase without disturbing the organic phase, and transfer the aqueous phase to a new 1.5 ml low adhesion microcentrifuge tube.
4. To the isolated aqueous phase, add 1/10 volume of 3 M NaCl, 1 μl 15 mg/ml GlycoBlue, and 3 volumes of 100 % ethanol. Mix thoroughly by inverting the tube at least five times and incubate at −20 °C or −80 °C for 20 min. Proceed to cDNA library preparation,

Stopping point: If kept at −20 °C, RNA can be safely stored for several months as an ethanol precipitate.

3.3 cDNA Library Preparation and Deep Sequencing

The following section describes the standard small RNA cDNA library preparation protocol described for cloning of small regulatory RNAs, found in ref. 16. Before generating the small cDNA libraries following the steps described below, we strongly recommend reading this protocol. The main differences in the procedure described here are: (a) the use of a non-barcoded 3′ adapter, (b) no spike-in of radioactive RNA size markers, and (c) no spike-in of calibrator oligoribonucleotides.

See **Note 16** for general guidelines for the cDNA library preparation.

3.3.1 3′ Adapter Ligation

1. Prepare 5′-^{32}P-labeled RNA size marker cocktail. Use of the size markers will control for successful ligation and indicate the length of the bands that need to be cut out of the gel.
   - Prepare a 20-well, 15 % denaturing polyacrylamide gel (15 cm wide, 17 cm long, 0.5 mm thick; 25 mL gel solution).
   - Pre-run the gel for 30 min at 30 W using 1× TBE buffer. While the gel is pre-running move on to point 3 of **step 1**.
   - Radiolabel the size markers individually in a 10 μl reaction containing 1 μM RNA, 10 U T4 polynucleotide kinase, and 50 μCi γ-^{32}P-ATP at 37 °C for 15 min
   - Quench the reactions from point 1 of **step 1** by adding 10 μl of denaturing formamide gel loading solution to each reaction.
   - Denature the RNA by incubating the tubes for 1 min at 90 °C.
   - Load each sample into one well of the 15 % denaturing polyacrylamide gel. In order to avoid cross-contamination, make sure to space the size markers with a minimum 2-well distance from each other.

- Run the gel for 50 min at 30 W using 1× TBE buffer, until the bromophenol blue dye is close to the bottom of the gel.
- Dismantle the gel, leaving it mounted on one glass plate. Using a scalpel or razor blade, cut crosses of approx. 1 cm length in three corners of the gel. Into these crosses, pipette 1 μl of the radioactive waste (stored in Subheading 3.2.3, **step 3**) to facilitate alignment of the gel to the phosphorimager paper printout. Wrap the gel in plastic film (e.g., Saran wrap) to avoid contamination of the phosphorimager screen.
- Expose the gel for 5 min to a phosphorimager screen at −20 °C.
- Align the gel on top of a printout scaled to 100 % according to the position of the three spots of radioactive waste. Cut out the radioactive bands corresponding to the length marker.
- Place the gel slices in 1.5 ml low adhesion microcentrifuge tubes and cover in 0.3 M NaCl (>300 μl). Elute the ligation product into the NaCl using constant agitation at 4 °C overnight (a rotating wheel works well).
- The following day, take off the supernatant, add 1 μl 15 mg/ml GlycoBlue, mix well, and follow with addition of 3 volumes of 100 % ethanol. Mix thoroughly by inverting the tube at least five times and incubate at −20 °C or −80 °C for 20 min.
- Centrifuge the precipitated RNA at >14,000 × *g* at 4 °C for 30 min.
- Remove the supernatant completely without disturbing the pellets. Air-dry the pellets for 10 min.
- Resuspend the pellets in 10 μl water and combine the solutions to obtain the concentrated size marker cocktail.
- Transfer 1 μl of this cocktail to a new low adhesion tube and dilute it 1:50 in water to obtain the diluted size marker cocktail. Mix by pipetting up and down several times.
- 10 μl of this diluted size marker cocktail will be used in point 3 of **step 1**. Store the remaining diluted and concentrated size marker cocktail at −20 °C for future PAR-CLIPs. One preparation of concentrated size marker cocktail can be used for multiple experiments. When diluting the size marker cocktail in future experiments take into account the 14.5 day half life of ^{32}P.

2. Spin sample from Subheading 3.2.5, **step 4** at >14,000 × *g* at 4 °C for 20 min. A blue pellet should be visible at the bottom of the tube.

3. Remove the supernatant completely without disturbing the pellet. Air-dry the pellet for 10 min.
4. Resuspend the pellet in 10 μl water.
5. Prepare the following reaction mixture for ligation of the adenylated 3′ adapter, multiplying the volumes by number of ligation reactions to be performed plus 2 extra volumes to include the diluted radioactive RNA size marker cocktail and to account for pipetting error:
   - 2 μl 10× RNA ligase buffer without ATP
   - 6 μl 50 % aqueous DMSO
   - 1 μl 100 μM adenylated 3′ adapter oligonucleotide
6. Add 9 μl of the reaction mixture to each sample, including the 10 μl of diluted radioactive RNA size marker cocktail.
7. Denature the RNA by incubating the tubes for 1 min at 90 °C. Immediately place the tubes on ice and incubate for 2 min.
8. Add 1 μl of Rnl2(1-249)K227Q ligase (1 μg/μl), swirl gently with your pipette tip, and incubate the tubes overnight on ice at 4 °C
9. Prepare a 20-well, 15 % denaturing polyacrylamide gel (15 cm wide, 17 cm long, 0.5 mm thick; 25 mL gel solution).
10. Pre-run the gel for 30 min at 30 W using 1× TBE buffer.
11. Add 20 μL of formamide gel loading solution to each 3′ adapter ligation reaction.
12. Denature the RNA by incubating the tubes for 1 min at 90 °C.
13. Load each sample into one well of the 15 % denaturing polyacrylamide gel. In order to avoid cross-contamination, make sure to space different samples appropriately; we recommend a two well distance.
14. Split the marker reaction, loading one half on opposite ends of the gel to frame the PAR-CLIP samples. Once again, avoid cross-contamination by keeping a two-well distance between samples and markers.
15. Run the gel for 45 min at 30 W using 1× TBE buffer, until the bromophenol blue dye is close to the bottom of the gel.
16. Dismantle the gel, leaving it mounted on one glass plate. Using a scalpel or razor blade, cut crosses of approximately 1 cm length in three corners of the gel. Into these crosses, pipette 1 μl of the radioactive waste (stored in Subheading 3.2.3, **step 3**) to facilitate alignment of the gel to the phosphorimager paper printout. Wrap the gel in plastic film (e.g., Saran wrap) to avoid contamination of the phosphorimager screen.
17. Expose the gel for at least 1 h to a phosphorimager screen, keeping the cassette at −20 °C to prevent diffusion of RNA.

If the radioactivity of the recovered RNA is weak, you can expose the gel overnight at −20 °C.

18. Align the gel on top of a printout scaled to 100 % according to the position of the three spots of radioactive waste. The 3′-ligated 19 and 35 nt markers should be visible on the printout, possibly with two additional lower bands representing unligated 19 and 35 nt marker. Using the ligated markers as guides, cut out sample RNA of 19–35 nt length, ligated to the 3′ adapter. Cut out the ligated markers as well (*see* **Note 17**).
19. Place the gel slices in separate 1.5 ml low adhesion microcentrifuge tubes and cover in 0.3 M NaCl (>300 μl). Elute the ligation product into the NaCl using constant agitation at 4 °C overnight (a rotating wheel works well).
20. The following day, take off the supernatant, add 1 μl 15 mg/ml GlycoBlue, mix well, and follow with addition of 3 volumes of 100 % ethanol. Mix thoroughly by inverting the tube at least five times and incubate at −20 °C or −80 °C for 20 min.

Stopping point: If kept at −20 °C, RNA can be safely stored for several months as an ethanol precipitate.

3.3.2 5′ Adapter Ligation

1. Centrifuge the precipitated RNA at >14,000 × *g* at 4 °C for 30 min. A blue pellet should be visible at the bottom of the tubes.
2. Remove the supernatant completely without disturbing the pellet. Air-dry the pellet for 10 min.
3. Resuspend the pellet in 9 μl water.
4. Prepare the following reaction mixture for ligation of the 5′ adapter, multiplying the volumes by number of ligation reactions to be performed plus 2 extra volumes to include the RNA size markers and to account for pipetting errors:
   - 2 μl 10× RNA ligase buffer with ATP
   - 6 μl 50 % aqueous DMSO
   - 1 μl 100 μM 5′ adapter oligonucleotide
5. Add 9 μl of the reaction mixture to each sample, including the 3′-ligated radioactive RNA size markers.
6. Denature the RNA by incubating the tubes for 1 min at 90 °C. Immediately place the tubes on ice and incubate for 2 min.
7. Add 2 μl T4 RNA ligase, swirl gently with your pipette tip, and incubate for 1 h at 37 °C. While the samples are incubating, prepare the polyacrylamide gel.
8. Prepare a 20-well, 12 % denaturing polyacrylamide gel (15 cm wide, 17 cm long, 0.5 mm thick; 25 mL gel solution).
9. Pre-run the gel for 30 min at 30 W using 1× TBE buffer.

10. Add 20 μL of formamide gel loading solution to each 5′ adapter ligation reaction.
11. Denature the RNA by incubating the tubes for 1 min at 90 °C.
12. Load the gel as described in Subheading 3.3.1, **steps 10** and **11**, and run for 45 min at 30 W using 1× TBE buffer, until the bromophenol blue dye is close to the bottom of the gel.
13. Image the gel as described in Subheading 3.3.1, **steps 13** and **14**, and excise the new ligation product (*see* **Note 18**).
14. Place the gel slices in 1.5 ml low adhesion microcentrifuge tubes and cover in 0.3 M NaCl (>300 μl). Elute the ligation product into the NaCl using constant agitation at 4 °C overnight (a rotating wheel works well).
15. The following day, take off the supernatant, add 1 μl 15 mg/ml GlycoBlue, mix well, and follow with addition of 3 volumes of 100 % ethanol. Mix thoroughly by inverting the tube at least five times and incubate at −20 °C or −80 °C for 20 min.

Stopping point: If kept at −20 °C, RNA can be safely stored for several months as an ethanol precipitate.

3.3.3 Reverse Transcription

1. Centrifuge the precipitated RNA at >14,000×*g* at 4 °C for 30 min. A blue pellet should be visible at the bottom of the tubes.
2. Remove the supernatant completely without disturbing the pellet and allow the pellet to air-dry for 10 min.
3. Resuspend the pellet in 4.6 μl water and transfer to a thermocycler tube.
4. Prepare the following reaction mixture for reverse transcription, multiplying the volumes by number of reverse transcription reactions to be performed plus 1 extra volume to account for pipetting errors:
   - 1.5 μl 0.1 M DTT
   - 3 μl 5× first-strand synthesis buffer
   - 4.2 μl 10× dNTPs
   - 1 μl 100 μM 3′ primer
5. Before addition of the reaction mixture, denature the RNA by incubating the tubes for 30 s at 90 °C in a thermocycler, and then hold at 50 °C.
6. Add 9.7 μL of the reaction mix to each sample and incubate for 3 min at 50 °C. Add 0.75 μl of Superscript III reverse transcriptase, mix gently by flicking the tube twice and incubate for 2 h at 50 °C.
7. Add 85 μl water and mix well.

Stopping point: cDNA can be stored indefinitely at −20 °C.

3.3.4 PCR Amplification

1. Prepare the following mix multiplied by the number of samples:
   - 40 µl 10× PCR buffer
   - 40 µl 10× dNTPs
   - 2 µl 100 µM 5′ primer
   - 2 µl 100 µM 3′ primer
   - 272 µl water
   - 89 µl of the reaction mix will be used in a pilot PCR reaction to determine the optimal number of PCR cycles for amplification, and the remaining mixture will be used for a large scale PCR.
2. To 89 µl of the reaction mix add 10 µl from the cDNA solution and 1 µl of Taq polymerase.
3. Perform a standard 100 µl, 30 cycle PCR with the following conditions: 45 s at 94 °C, 85 s at 50 °C, 60 s at 72 °C.
4. Beginning with the 12th cycle and ending with the 30th cycle, remove a 10 µl aliquot from each PCR reaction every 3 cycles (i.e., at cycles 12, 15, 18, etc.).
5. Analyze the 10 µl aliquots on a 2.5 % agarose gel alongside a 25 bp ladder. The expected PCR product should appear between 95 and 110 bp. When ligated and amplified with the correct primers, the 19 and 35 nt markers appear at 95 and 110 bp respectively. Often, a lower band appears at 72 bp corresponding to the direct ligation products of the 3′ and 5′ adapters. Define the optimal cycle number for cDNA amplification, which should be within the exponential amplification phase of the PCR, approximately 5 cycles away from reaching the saturation level of PCR amplification (*see* **Note 19**).
6. Using the remaining PCR cocktail, perform three 100 µl PCR reactions with the optimal cycle number identified above.
7. Combine the individual 100 µl reactions and precipitate with 3 volumes of 100 % ethanol.
8. Take up the pellet in 60 µl 1× DNA loading dye.
9. Run the sample on two wells of a 2.5 % agarose gel alongside a 25 bp ladder.
10. Visualize the DNA on a UV transilluminator and excise the gel piece containing cDNA between 85 and 120 bp of length.
11. Extract the DNA using the Qiagen MinElute Gel Extraction Kit, following the instructions of the manufacturer. Use 30 µl elution buffer to recover the DNA.
12. Submit 10 µl of the purified cDNA to Illumina sequencing. We recommend using 50 cycle single-end sequencing on a HiSeq machine.

3.3.5 Optional: Determination of Incorporation Levels of 4-Thiouridine into Total RNA

1. Supplement growth medium with 100 μM of 4SU 16 h prior to harvest, provide regular media to one control plate.
2. The following day, harvest cells using a cell scraper and spin down at 500 × *g* for 5 min at 4 °C.
3. Remove supernatant and resuspend the pellet in 3 volumes of TrIzol reagent (Sigma), follow the manufacturer's instructions.
4. Further purify total RNA using Qiagen RNAeasy according to the manufacturer's instructions (*see* **Note 20**).
5. Digest and dephosphorylate total RNA to single nucleosides by incubating 40 μg of purified total RNA for 16 h at 37 °C with 0.4 U bacterial alkaline phosphatase (e.g., Worthington Biochemical) and 0.09 U snake venom phosphodiesterase (e.g., Worthington Biochemical) in a 30 μl volume.
6. As a reference standard, use a synthetic 4SU labeled RNA (previously we used CGUACGCGGAAUACUUCGA(4SU)U), which is subjected to complete enzymatic digestion.
7. Separate the resulting mixtures of ribonucleosides by HPLC on a Supelco Discovery C18 (bonded phase silica 5 μM particle, 250 × 4.6 mm) reverse phase column (Bellefont). HPLC buffers are 0.1 M triethylammonium acetate (TEAA) in 3 % acetonitrile (A) and 90 % acetonitrile in 0.1 M TEAA (B).
8. Use an isocratic gradient: 0 % B for 15 min, 0–10 % B for 20 min, 10–100 % B for 30 min.
9. Clean HPLC column with a 5 min 100 % wash between runs.

3.4 PAR-CLIP Analysis

With current depths of Illumina sequencing reaching >200 million sequence reads per sample, PAR-CLIP data analysis requires sophisticated approaches to identify binding sites [17]. Several biocomputational pipelines for PAR-CLIP data analysis have been made available, including PARalyzer [18], PIPE-CLIP [19], WavclusteR [20], doRina [21], CLIPZ [22], Starbase [23], miRTarCLIP [24], Piranha [25], and dCLIP [26]. After initial analysis, you may calculate the common sequence motifs of the RRE using one of the several programs initially developed for the analysis of transcription-factor binding sites on DNA, including MEME [27], MDScan [28], PhyloGibbs [29], cERMIT [30], and Gimsan [31].

Generally, the analysis of the sequence reads begins by alignment to the genome, allowing for at least one error (substitution, insertion, or deletion) to capture cross-linked reads with cross-linking-induced mutations. Next, overlapping sequence reads are grouped, taking into account the frequency of cross-linking-induced mutations. To allow insights into the RBP's binding preferences, these groups of overlapping sequence reads can then be mapped against the transcriptome to annotate and categorize them as derived from 5′ untranslated region (UTR), coding sequence

(CDS), 3′UTR, introns, rRNA, long noncoding RNAs, tRNAs, and so forth.

The frequency of the T-to-C mutations (or G-to-A mutations when using 6SG) allows ranking of groups to predict those interactions with the highest functional impact. In addition, it may be useful to provide a limited set of high-confidence interaction sites as input into motif-finding programs to facilitate the detection of the underlying RRE. Some of the analysis pipelines, such as PARalyzer, take advantage of the frequency and distribution of cross-linking-induced mutations to predict the shortest possible region of interaction between RBP and RNA that harbors the RRE.

CLIP-based approaches provide a genome-wide view of the protein–RNA interaction sites and routinely identify tens of thousands of interaction sites in the transcriptome. However, additional experimentation—as well as clear ranking of binding sites—is necessary to relate RNA binding to phenotypes arising from knockout, overexpression, or mutation of the RBP. For example, the effect of RNA binding on transcript stability and alternative splicing can be assayed using microarray analysis and RNA sequencing analysis. Quantitative proteomics (SILAC, iTRAQ) and ribosome profiling are increasingly available as methods to assess translational regulation by RBPs [32]. Analysis of RBPs involved in RNA transport and other processes may require the development of more specialized assays.

4 Notes

1. For NP40 Lysis Buffer prepare a stock of 5× buffer without DTT and protease inhibitors. Add DTT and protease inhibitor to 1× buffer directly before use.
2. Not every antibody will retain its binding ability in 500 mM KCl—adjust the salt concentration accordingly for the high salt wash buffer. If in doubt use lysis buffer for washing. Also add DTT and protease inhibitor directly to high salt wash buffer before experiment.
3. This protocol describes the procedure for analysis of endogenously expressed, or recombinant constitutively expressed, or inducibly expressed RBPs. The PAR-CLIP protocol will work with any cell line expressing detectable levels of RBP as long as there is an efficient antibody for *immunoprecipitation* (IP). However, some antibody quality testing is necessary before beginning PAR-CLIP. If using an antibody that specifically recognizes your RBP-of-interest, perform stringent quality testing of the IP with your antibody *before* attempting PAR-CLIP. We recommend transiently transfecting the cells with a vector for expression of the protein of interest with an

N-terminal or C-terminal epitope fusion, such as FLAG, MYC, or HA. Follow transfection with IP using an RBP-specific antibody and Western blotting for the epitope tag using reliable commercial antibodies. At this time you should also test for the maximal monovalent salt concentration compatible with your IP. Increasing salt concentration will result in fewer co-purifying proteins and RNAs but can also lead to loss of bound RNP. For reference, the FLAG antibody tolerates up to 500 mM KCl.

4. Guidelines for the use of *4-thiouridine* (4SU) may need to be adapted for use in the desired cell lines or model organisms; the concentration of 4SU and the length of UV-light exposure in this protocol were optimized for HEK293 cells. For other cell lines, the user may want to determine the optimal, non-toxic 4SU concentration and labeling time. In cell lines or model organisms with weak 4SU uptake, it may be necessary to enhance or introduce expression of nucleoside transporters, such as *uracil phophoribosyltransferase* (UPRT) [33, 34]. We have also included an optional section at the end of the PAR-CLIP procedure for determining the incorporation of 4SU into total RNA. The energy dose of UV light necessary for cross-linking may vary due to differing transparency of the sample compared to mammalian cells grown in monolayers. For example, cells plated as dense suspensions, yeast, and worms exhibit higher opacity [35].
5. We recommend use of positive and negative controls, particularly when performing the pilot PAR-CLIP experiments. An appropriate negative control could comprise the use of IgG isotype control as a substitute for the RBP antibody; this will allow the user to visualize fragments of abundant cellular RNAs, as well as RNPs co-purifying through nonspecific interactions with antibodies and magnetic beads (Thermo Scientific MA1-10407). For a positive control, plasmids encoding FLAG/HA-tagged RBPs previously characterized by PAR-CLIP are available on www.addgene.org.
6. The on-bead RNase T1 digestion described in Subheading 3.2.2, **step 7** should be optimized for your individual RBP. Each RBP binding footprint provides a different level of protection from RNase T1, resulting in shorter or longer RNA fragments after the RNA is isolated by Proteinase K digestion. RNA fragments between 19 and 35 nt are ideal for small cDNA library preparation. Fragments shorter than 19 nt have a higher probability of mismapping compared to longer reads. Fragments longer than 35 nt cannot be fully sequenced by standard 50 base single-end sequencing. The concentration of RNase T1 suggested in this protocol may be too high for certain RBPs, resulting in RNA less than 19 nt long. To determine the correct

concentration of RNase T1 for the on-bead RNase T1 digestion, perform the PAR-CLIP protocol through Subheading 2. When you reach Subheading 3.2.2, **step 7**, perform a set of separate digestions with RNase T1 concentrations ranging from 0 to 100 U/μl. After Subheading 2 is complete, analyze the resulting RNA on a denaturing polyacrylamide gel as described in Subheading 3.3.1, **steps 10–16**. If the majority of RNA is below 19 nt, over-digestion has occurred and the RNase T1 concentration must be reduced appropriately.

7. Use low adhesion microcentrifuge tubes (e.g., siliconized or DNA LoBind tubes) for all manipulations of the small RNAs after the Proteinase K digestion. The minute amounts of small RNAs to be recovered after RNA isolation will readily adsorb to the walls of standard tubes.
8. It is also possible to use 100 μM of 6-thioguanine (6SG) as the photoactivatable ribonucleoside. 6SG has a lower cross-linking efficiency compared to 4SU and will result in a G-to-A mutation instead of a T-to-C mutation at the cross-linking site.
9. A thin film of remaining media helps prevent cells from drying, and does not interfere with cross-linking.
10. Incubation at room temperature is also sufficient if there are no means of incubation at 22 °C.
11. Guidelines for working with magnetic beads:
    - Before pipetting beads from the source container, always mix thoroughly by shaking or vortexing.
    - To prevent drying and loss of function, do not leave beads uncovered for prolonged periods of time

Step-by-step for washing magnetic beads:

1. Place the beads in suspension on a magnetic separator and let stand for 1 min or until solution clears.
2. Carefully remove buffer from the tube without disturbing the beads.
3. Add buffer to the tube while the tube is on the magnetic separator.
4. Remove the tube from the magnetic separator and resuspend the beads either by flicking, shaking, or vortexing. To prevent loss of beads, we do not recommend mixing by pipetting.
5. Briefly centrifuge the tube to collect beads caught on the tube cap.
6. Place the beads in suspension on a magnetic separator and let stand for 1 min or until solution clears.

7. Remove the supernatant and resuspend in the appropriate buffer, or repeat **steps 1–5** for additional wash steps.
12. For small volumes of lysate do not use less than 35 μl of magnetic beads to account for minor loss during handling.
13. Make sure that you do not exceed the maximum salt concentration at which the antibody recognizes its antigen.
14. Optional: Reduce IP background by performing a high salt wash. Replace 1× NP40 lysis buffer with 1× high salt wash buffer. Only perform this step if you are confident you will not exceed the maximum salt concentration at which the antibody recognizes its antigen.
15. To avoid bead damage, do not expose magnetic beads to high DTT concentration for prolonged time.
16. Take care to avoid contamination of the minute amounts of RNA, e.g., with RNases. Use RNAse-free water and store RNA at −20 or −80 °C. Previously prepared cDNA libraries may contaminate lab surfaces and equipment and will readily amplify during PCR. Use sterile filter tips wherever possible.
17. Avoid recovering RNA <19 nt, which have the potential to complicate the subsequent bioinformatics analysis, as they have a higher probability of mismapping compared to longer reads.
18. Optional: Excise the markers and keep them as controls for the reverse transcription and PCR.
19. As the PCR reaction approaches saturation of PCR product, reagents within the reaction become limiting, leading to selective amplification of certain transcripts over others.
20. It is important to add 0.1 mM dithiothreitol (DTT) to wash buffers and subsequent enzymatic steps to prevent oxidization of 4SU during RNA isolation and analysis.

References

1. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136:731–745
2. Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 136:688–700
3. Martin KC, Ephrussi A (2009) mRNA localization: gene expression in the spatial dimension. Cell 136:719–730
4. Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157:77–94
5. Lukong KE, Chang K-W, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends Genet 24:416–425
6. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793
7. Castello A, Fischer B, Hentze MW, Preiss T (2013) RNA-binding proteins in Mendelian disease. Trends Genet 29:318–327
8. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
9. König J, Zarnack K, Luscombe NM, Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. Nat Publ Group 13:77–83
10. Tenenbaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci U S A 97:14085–14090

11. Gilbert C, Svejstrup JQ (2006) RNA immunoprecipitation for determining RNA-protein associations in vivo. Curr Protoc Mol Biol Chapter 27, Unit 27.4–27.4.11
12. Gerber AP, Luschnig S, Krasnow MA et al (2006) Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. Proc Natl Acad Sci U S A 103:4487–4492
13. López de Silanes I, Zhan M, Lal A et al (2004) Identification of a target RNA motif for RNA-binding protein HuR. Proc Natl Acad Sci U S A 101:2987–2992
14. Maes OC, Chertkow HM, Wang E, Schipper HM (2009) MicroRNA: implications for Alzheimer disease and other human CNS disorders. Curr Genomics 10:154–168
15. Hafner M, Landthaler M, Burger L et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
16. Hafner M, Renwick N, Farazi TA et al (2012) Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. Methods 58:164–170
17. Ascano M, Hafner M, Cekan P et al (2011) Identification of RNA-protein interaction networks using PAR-CLIP. WIREs Interdiscip Rev RNA 3:159–177
18. Corcorabin DL, Georgiev S, Mukherjee N et al (2011) PARalyzer: definition of RNAding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79
19. Chen B, Yun J, Kim MS et al (2014) PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. Genome Biol 15:1–10
20. Sievers C, Schlumpf T, Sawarkar R, Comoglio F, Paro R (2012) Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. Nucleic Acids Res 40(2):160
21. Anders G, Mackowiak SD, Jens M et al (2012) doRiNA: a database of RNA interactions in post-transcriptional regulation. Nucleic Acids Res 40:D180–D186
22. Khorshid M, Rodak C, Zavolan M (2011) CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic Acids Res 39:D245–D252
23. Yang JH, Li JH, Shao P et al (2011) starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res 39:D202–D209
24. Chou CH, Lin FM, Chou MT et al (2013) A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. BMC Genomics 14(Suppl 1):S2
25. Uren PJ, Bahrami-Samani E, Burns SC et al (2012) Site identification in high-throughput RNA-protein interaction data. Bioinformatics 28:3013–3020
26. Wang T, Xie Y, Xiao G (2014) dCLIP: a computational approach for comparative CLIP-seq analyses. Genome Biol 15:R11
27. Bailey TL (2002) Discovering novel sequence motifs with MEME. Curr Protoc Bioinformatics Chapter 2, Unit 2.4–2.4.35
28. Liu XS, Brutlag DL, Liu JS (2002) An algorithm for finding protein–DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol 20:835–839
29. Siddharthan R, Siggia ED, van Nimwegen E (2005) PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. PLoS Comput Biol 1:e67
30. Georgiev S, Boyle AP, Jayasurya K et al (2010) Evidence-ranked motif identification. Genome Biol 11:R19
31. Ng P, Keich U (2008) GIMSAN: a Gibbs motif finder with significance analysis. Bioinformatics 24:2256–2257
32. Brewis IA, Brennan P (2010) Proteomics technologies for the global identification and quantification of proteins. Adv Protein Chem Struct Biol 80:1–44
33. Guruharsha KG, Rual JF, Zhai B et al (2011) A protein complex network of Drosophila melanogaster. Cell 147:690–703
34. Kucerova L, Poturnajova M, Tyciakova S, Matuskova M (2012) Increased proliferation and chemosensitivity of human mesenchymal stromal cells expressing fusion yeast cytosine deaminase. Stem Cell Res 8:247–258
35. Jungkamp AC, Stoeckius M, Mecenas D et al (2011) In vivo and transcriptome-wide identification of RNA binding protein target sites. Mol Cell 44:828–840

Chapter 11

Profiling the Binding Sites of RNA-Binding Proteins with Nucleotide Resolution Using iCLIP

FX Reymond Sutandy, Andrea Hildebrandt, and Julian König

Abstract

The importance of posttranscriptional regulation in cellular metabolism has recently gone beyond what was previously appreciated. The regulatory mechanisms are controlled by RNA-binding proteins (RBPs), which form complexes with RNA and regulate RNA processing, stability, and localization, among others. Consistently, mutations in RBPs result in defects in developmental processes, diseases, and cancer. Gaining deeper insights into the biology of RNA–RBP interactions will lead to a better understanding of regulatory processes and disease development. Several techniques have been developed to capture the properties of RNA–RBP interactions. Furthermore, the development of high-throughput sequencing has broadened the capability of these methods. Here, we summarize individual-nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP), a powerful technique that provides genome-wide information on RNA–RBP interactions at nucleotide resolution. In this chapter, we outline the iCLIP protocol and list possible controls that allow a targeted and cost-minimizing optimization of the protocol for an RBP-of-interest. Moreover, we provide notes on experimental design and a troubleshooting guideline for common problems that can occur during iCLIP library preparation.

Key words RNA, RNA-binding protein, iCLIP, RBP binding site, Protein–RNA interactions

1 Introduction

RNA processing steps, such as splicing and polyadenylation, are major regulatory mechanisms in posttranscriptional gene regulation. Other points of action include RNA stability and localization by which the cell can coordinate RNA turnover and local protein synthesis. During all these processes, RNA-binding proteins (RBPs) act as the minders and controllers.

RNA–protein interactions have been subject to basic research for more than 40 years and are still investigated to date [1, 2]. As early as 1979, Jay R. Greenberg found that cross-linking of proteins to RNA is feasible at a wavelength of 254 nm. This led to the conclusion that RNA and proteins can be found in close proximity in the cell [3]. Many RBPs were originally described by their

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_11,

conserved RNA binding domains, which exist in single or multiple copies and enable the RBPs to associate with RNA in a sequence- or structure-dependent manner [4]. More recently, approaches using protein microarrays and proteomics identified several hundred novel RBPs in yeast [5, 6], and further screens for human RBPs based on oligo(dT)-pulldowns coupled to mass spectrometry added hundreds of novel RBPs to the list [7–9]. Notably, in these screens, an RNA binding function was often assigned to proteins that had previously only been known to be involved in other cellular processes, such as metabolic functions.

Following from their abundance and widespread function, mutations in RBPs can disturb fundamental biological processes and are often linked to human diseases and cancer [10]. For example, loss of function of the fragile X mental retardation protein (FMRP) leads to the fragile X syndrome, a form of mental retardation [11]. Therefore, identifying RNA targets and especially disease-associated RBP-RNA interactions will be beneficial to understand the molecular mechanisms of disease and to ultimately develop treatments.

In order to identify the RNA targets of an RBP-of-interest, RNA can be cross-linked to directly interacting proteins in vivo using UV irradiation [3]. Cross-linking is vital, since earlier approaches have shown that RNA–RBP complexes can reassociate during affinity purification [12]. The method UV cross-linking and immunoprecipitation (CLIP) exploits UV cross-linking to obtain covalently bound RBP-RNA complexes, which are immunoprecipitated with specific antibodies against the RBP-of-interest [13, 14]. It is then possible to extract the interacting RNAs in subsequent steps. In a related method called PAR-CLIP, the cross-linking is achieved through the use of photoreactive nucleosides, such as 4-thiouridine, which can be specifically cross-linked at 365 nm [15]. Coupled with high-throughput sequencing (HITS-CLIP or CLIP-Seq) [16], CLIP allows to map RNA–RBP interactions on a genomic scale, which can be used for instance to identify target sequences [17] or to compute RNA-maps [16].

We recently introduced individual-nucleotide resolution CLIP (iCLIP), which allows to quantitatively map RNA–RBP interaction sites with high resolution and efficiency [18]. Basis to the iCLIP procedure is the fact that during cDNA synthesis from the co-purified RNAs, the reverse transcriptase often truncates at the cross-link site, at which a small residual of the RBP is left after protein digestion [19]. The iCLIP library preparation protocol allows to capture these truncated cDNAs, which provide a footprint of the RBP binding site [20]. More precisely, the cDNA sequences start one nucleotide downstream of the cross-link sites, thereby providing nucleotide-resolution information on RBP binding. To date, iCLIP has been widely applied to investigate diverse functions of RBPs in RNA metabolism, such as ,regulating alternative

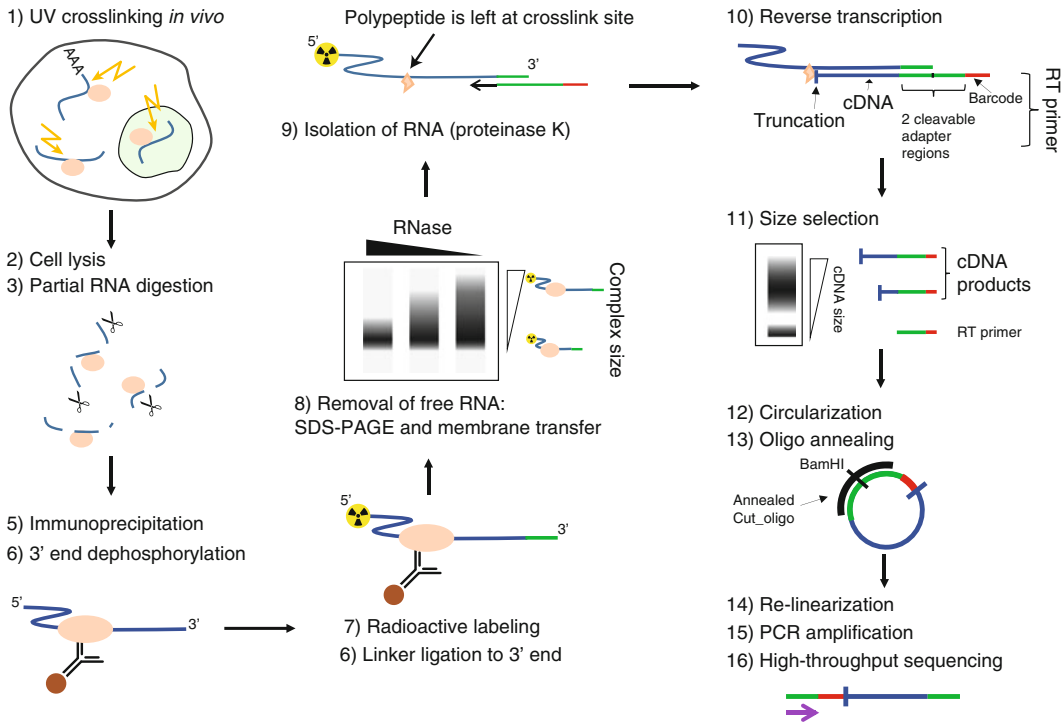

Fig. 1 Schematic overview of the iCLIP protocol. The experiment starts with UV irradiation of the samples, followed by cell lysis and partial RNase digestion. The specific protein–RNA complexes are then immunoprecipitated with a specific antibody, and an adaptor is ligated to the co-purified RNAs. The complexes are subsequently run on an SDS-PAGE and transferred to a nitrocellulose membrane, which will remove most unspecifically bound RNAs. Proteinase K digestion is applied to extract the cross-linked RNAs, which are then subjected to reverse transcription. The resulting cDNAs are size-selected and circularized. The following linearization step provides the basis for amplification of the iCLIP library, which is then ready for high-throughput sequencing

splicing and RNA stability [21, 22], mediating RNA modifications [23], or binding small RNAs [24, 25], among others.

In this chapter, we introduce the iCLIP protocol, including a description of essential controls and guidelines for troubleshooting. The iCLIP experiment starts with UV irradiation of the sample (Fig. 1), which introduces a covalent bond between the RBP and the interacting RNAs. The cells are then lysed, and a partial RNase digestion is applied to cut the interacting RNAs down to a specific size range, which is optimal for library preparation. An example of isolated RNAs under different RNase conditions is shown in Fig. 2b. This is followed by an immunoprecipitation step to capture the RBP-of-interest in complex with the cross-linked RNA fragments. Due to the covalent bond between the RBP and the RNA fragments, a stringent purification and washing scheme can be applied to remove unspecific interactions. Next, the 3′ ends of the RNA fragments are dephosphorylated and then ligated to a DNA linker, while the 5′ ends are radioactively labeled to enable

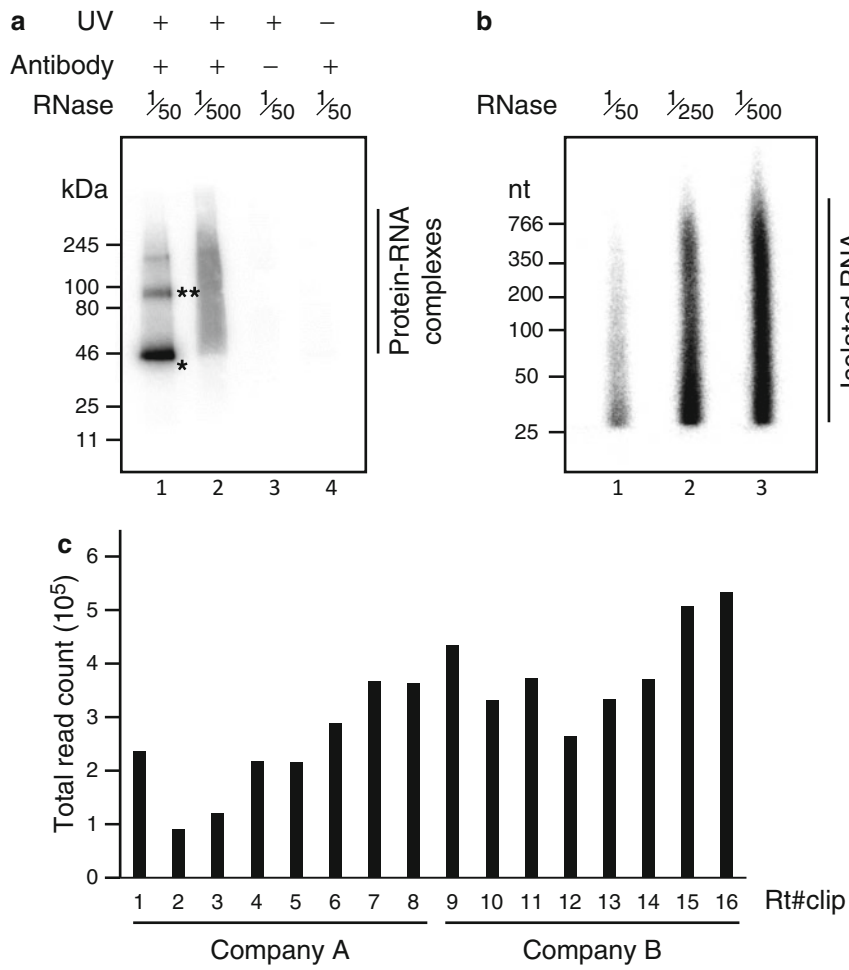

Fig. 2 Visualization of RNA–RBP complexes and isolated RNAs. (**a**) Radioactively labeled RNA–RBP complexes upon membrane transfer. *Lane 1*, high RNase concentration resulting in short RNA fragments on the RBP. The radioactive signal appears immediately upstream of the estimated size of the RBP (labeled by *). **Higher molecular weight band, indicating two proteins cross-linked to the same RNA molecule. *Lane 2*, low RNase concentration producing a smear of RNA–RBP complexes with RNA fragments of varying sizes. *Lane 3*, no-antibody control. *Lane 4*, no-UV cross-linking control. (**b**) Gel separation of RNA fragments that were isolated from the nitrocellulose membrane after digestion with different RNase concentrations. In this experiment, the 1:500 dilution shows the highest amount of RNAs in the optimal size range of 30–150 nt. (**c**) Comparison of read numbers obtained from the same sample that was processed with Rt#clip primers with different barcodes and synthesized by two different companies. The variability in library preparation efficiency needs to be considered when comparing between samples

their visualization on an autoradiograph upon SDS-PAGE and transfer to a nitrocellulose membrane (Fig. 2a). The latter is critical to further remove non-cross-linked RNAs from the sample. The cross-linked RNAs are then extracted from the membrane by proteinase K digestion and subjected to reverse transcription (RT) with RT primers complementary to the DNA linker sequence. During this step, most cDNAs will truncate at the cross-link site,

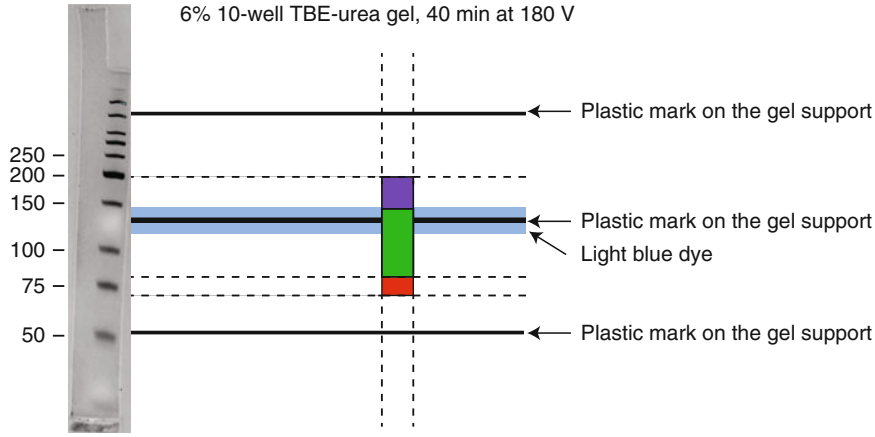

Fig. 3 Gel cutting mask for cDNA size selection. This mask can be used to align the cDNA gel via the blue dye as well as the rims on the plastic gel support on the 6 % TBE-urea pre-cast gel. The size marker should be used as an indicator that the gel has indeed run properly. The mask indicates three areas for cutting: *Red*, cDNA size ±70–80, insert size ±18–28: This band contains high cDNA complexity (i.e., high concentration of cDNA), but due to the short insert sizes, not all cDNAs can be mapped as unique hits to the genome. Also, this band has a tendency to isolate the primer contaminant. We recommend to use this band only when binding to short RNAs, such as miRNAs, is of interest. *Green*, cDNA size ±80–150 nt, insert size ±28–98: This band contains the best cDNAs that are long enough to map as unique hits to the genome. It should not have the primer contaminant and the complexity is generally high. *Violet*, DNA size ±150–200, insert size ±98–148: These cDNAs will all map to the genome, but cDNA complexity is usually limited. We nevertheless recommend to include this band initially for testing

where a small peptide residue of the RBP remains after proteinase K digestion. Importantly, the RT primers include a random barcode sequence, which facilitates removal of PCR duplicates during computational data analysis, allowing to ultimately keep only unique RT products. Gel purification and size selection of the cDNAs remove residual RT primers. Figure 3 provides a mask for selecting cDNAs in the optimal size range from these gels. Upon cDNA circularization, an oligonucleotide is annealed to the linker region to create a double-stranded cutting site for a restriction enzyme. This re-linearization renders cDNAs that are flanked on either side by adapter regions for PCR amplification. After another round of size selection to remove excess PCR primers, the iCLIP library is ready for high-throughput sequencing.

Positive and negative controls are critical during the iCLIP library preparation. Table 1 summarizes important controls that should be used, especially when optimizing the procedure. Since iCLIP is an extensive protocol, these controls serve as early indicators for problems with library quality, which can reduce working time and costs and decrease the risk of producing suboptimal iCLIP libraries. Table 2 lists troubleshooting strategies for common problems in iCLIP library preparation.

Table 1

Possible controls during iCLIP library preparation

Type of control	Purpose
Negative control without UV cross-link	Confirm that the radioactive signal comes from cross-linked molecules. This is to exclude that for example the protein itself can be radioactively labeled through autophosphorylation or a contaminating protein kinase.
Negative control without antibody	Detect possible contaminations. Unspecific proteins binding to the beads would still give a signal in this control. To prevent this, washing steps could be carried out more stringently.
Negative control upon knock-down of the target protein	Evaluate the specificity of the antibody used in the immunoprecipitation. A specific antibody would produce weaker or no protein–RNA signals in the autoradiograph image.
Positive control with antibody against a different protein	Cross-linking and IP efficiency are protein-dependent. This control ensures that all steps of the protocols are working and can help to estimate whether the amount of cross-linked RNA is sufficient to successfully prepare an iCLIP library for the protein-of-interest. Best for this control are antibodies against proteins, which were successfully used in previous iCLIP experiments.
Visualization of RNA sizes using different RNase conditions	The RNase concentration has to be optimized to gain suitable RNA sizes. In order to determine the RNA size range, the RNA fragments resulting from different RNase concentrations can be visualized on a gel after their isolation from the membrane using proteinase K digestion (Fig. 2b).

2 Materials

2.1 Buffers

Store all buffers in the fridge and keep them on ice during experiments.

1. Lysis buffer: 50 mM Tris–HCl, pH 7.4, 100 mM NaCl, 1 % Igepal CA-630 (Sigma), 0.1 % SDS, 0.5 % sodium deoxycholate. On the day of the experiment, add 1/100 volume of protease inhibitor cocktail III (Calbiochem/Merck) to the amount of buffer required for lysis (but not washing) (*see* **Note 1**).
2. High-salt wash buffer: 50 mM Tris–HCl (pH 7.4), 1 M NaCl, 1 mM EDTA, 1 % Igepal CA-630, 0.1 % SDS, 0.5 % sodium deoxycholate.
3. PNK buffer: 20 mM Tris–HCl (pH 7.4), 10 mM $MgCl_2$, 0.2 % Tween-20.
4. 5× PNK pH 6.5 buffer: 350 mM Tris–HCl (pH 6.5), 50 mM $MgCl_2$, 5 mM dithiothreitol. Freeze aliquots of the buffer, do not thaw and freeze again.
5. 4× ligation buffer: 200 mM Tris–HCl (pH 7.8), 40 mM $MgCl_2$, 4 mM dithiothreitol. Freeze aliquots of the buffer, do not thaw and freeze again.

Table 2

Troubleshooting for common problems during iCLIP library preparation

Step	Problem	Possible reason	Solution
Autoradiograph imaging (Subheading 3.10)	No/faint band	Failure in immunoprecipitation	Optimize the antibody used in the immunoprecipitation step.
		Poor cross-linking efficiency	Low cross-linking efficiency would cause huge loss of RNA–protein complexes, which leads to poorly detectable bound RNAs in the autoradiograph. Since the cross-linking efficiency is protein-dependent, you could either increase irradiation time or try 4-thiouridine cross-linking at 365 nm.
	No/little smear	Too much RNase digestion	Adjust RNase concentration (Fig. 2b).
	Strong band in the no-UV cross-linking control	The protein itself is radioactively labeled (*see* also Table 1)	In principle, this does not interfere with the isolation of bound RNA fragments. However, the presence of this signal has to be considered when interpreting the signal strength of the cross-linked samples.
	Multiple bands	Protein contaminations	Apply more washing steps.
		Protein dimerization	Some proteins might form a dimer, either by themselves or as a result of the purification tag that is fused to the proteins. In this case, RNA from all bands should be extracted.
		Two proteins are cross-linked in close proximity on the same RNA	This phenomenon could result in a "supershift" at the added size of both proteins. In this case, RNA from all bands should be extracted.
RNA precipitation (Subheading 3.11)	No RNA pellet	Some residues, such as high salt, gel, or phenol	Emergency action: Do not take off the supernatant, but split the sample into two tubes. Add 200 µl TE buffer followed by 20 µl sodium acetate and 500 µl 100 % ethanol to each tube to further dilute the sample. Incubate overnight at −20 °C and spin the samples again on the next day.
Post-PCR (Subheading 3.16)	Unspecific bands above the expected size	Overamplification	Secondary PCR products can be removed by reducing the number of amplification cycles. The post-PCR gel size selection also helps to remove undesired PCR products before high-throughput sequencing.
	Broader band distribution	Suboptimal cDNA size selection	Optimize the cDNA gel size selection step (Subheading 3.13) or perform post-PCR gel size selection.

6. PK buffer: 100 mM Tris–HCl (pH 7.4), 50 mM NaCl, 10 mM EDTA.
7. PK buffer/7 M urea: 100 mM Tris–HCl (pH 7.4), 50 mM NaCl, 10 mM EDTA, 7 M urea.

2.2 Linker and Primer Sequences

1. Pre-adenylated linker L3-App, rApp/AGATCGGAAGAGCGGTTCAG/ddC/ (IDT)

   The linker is modified with adenylation at the 5′ end (rApp) and dideoxycytidine (ddC) at the 3′ end.

2. Each RT primer (Rt#clip) has a different experimental barcode, which provides the possibility for sample multiplexing during high-throughput sequencing. In addition, all Rt#clip primers contain a random barcode sequence (NNxxxxNNN, where N is any base and x the bases of the defined experimental barcode) for the purpose of duplicate removal during sequencing data analysis. Here are examples of some Rt#clip primers:

Rt1clip	X33/NNAACCNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt2clip	X33/NNACAANNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt3clip	X33/NNATTGNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt4clip	X33/NNCGCCNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt5clip	X33/NNGCCANNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt6clip	X33/NNGACTNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt7clip	X33/NNGTGGNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt8clip	X33/NNTATTNNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC
Rt9clip	X33/NNTTAANNNAGATCGGAAGAGCGTCGTG gatcCTGAACCGC

   X33 = 5′ phosphate (*see* **Note 2**)

3. Cut_oligo, GTTCAGGATCCACGACGCTCTTCaaaa (IDT)
4. For PCR amplification and high-throughput sequencing, we used the official Illumina P3 and P5 primers:

   P5Solexa: AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCT

   P3Solexa: CAAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCT

3 Methods

All steps in this protocol should be performed on ice unless stated otherwise.

3.1 UV Cross-Linking

The following iCLIP protocol uses cell culture samples.

1. Remove media and add 6 ml cold PBS to cells growing in a 10 cm plate, remove lid and place on ice (usually enough for three immunoprecipitations).
2. Irradiate once with 150 mJ/cm^2 in a Stratalinker 2400 at 254 nm (*see* **Note 3**).
3. Harvest cells by scraping.
4. Add 2 ml suspension to each microtube (RNase-free, nonstick), spin at 0.4 g for 1 min at 4 °C, and remove supernatant.
5. Snap-freeze pellets on dry ice and store at −80 °C until use (or use directly).

3.2 Bead Preparation (*See* Notes 4 *and* 5)

1. Add 50–100 µl of protein G Dynabeads (Life Technologies) per experiment to a microtube (*see* **Note 6**).
2. Wash beads 2× with lysis buffer.
3. Resuspend beads in 100 µl lysis buffer with 2–10 µg antibody per experiment (*see* **Note 7**).
4. Rotate tubes at room temperature for 30–60 min (until lysate is ready).
5. Wash 1× with high-salt wash buffer.
6. Wash 2× with lysis buffer and leave in 400 µl lysis buffer with added protease inhibitors (*see* Subheading 2.1) until ready to proceed.

3.3 Pellet Resuspension

When ready to proceed, resuspend the cell pellets (from Subheading 3.1) in 1 ml lysis buffer. This should result in a protein concentration of ~2 mg/ml (*see* **Note 8**). Do not forgot to add the proteinase inhibitor to the lysis buffer before use (*see* Subheading 2.1).

3.4 Sonication of Samples (Optional Step) (*See* Notes 9 *and* 10)

Variant A: Sonicate sample on ice. The probe should be approximately 0.5 cm from the bottom of the tube and not touching the tube sides in order to avoid foaming. Sonicate 2× with 10 s bursts at 5 dB. Clean the probe by sonicating water before and after sample treatment.

Variant B: Transfer sample to 1.5 ml microtubes and use Bioruptor for five cycles with alternating 30 s on/off at low intensity (needs to be switched on in advance to allow for cooling down of the water bath). Six samples can be sonicated at the same time.

3.5 Partial RNase Digestion

1. Make 1/500 RNase I (Ambion) dilution in lysis buffer and add 10 μl to the lysate together with 2 μl Turbo DNase (Ambion) (*see* **Note 11**).

   High-RNase control (*optional*, recommended for initial optimizations): Treat one sample with high RNase: prepare a 1/50 RNase I dilution in lysis buffer and add 10 μl to the lysate together with 2 μl Turbo DNase. To minimize the use of reagents, it is possible to use only 1/5 of the cell lysate and all other reagents for this experiment (*see* **Notes 12** and **13**). Since this RNA is too short for DNA linker ligation, this control can skip 3′ end dephosphorylation and linker ligation and go straight from Subheadings 3.6–3.9.
2. Incubate for 3 min at 37 °C shaking at 1100 rpm in thermomixer. After incubation transfer to ice for 3 min (*see* **Note 14**).
3. Spin at 4 °C at top speed for 10 min and transfer the supernatant to a new 1.5 ml microtube.
4. Load 500 μlof the lysate onto a Proteus Clarification Mini Spin Column (Generon). Spin at 4 °C at 16,000 × *g* for 1 min. Transfer flow-through to a new microtube. Repeat with second half of the lysate and combine both (*see* **Note 15**).

3.6 Immunoprecipitation

1. Add the cell extract to the beads.
2. Rotate bead/lysate mix for 1 h (max. 2 h) at 4 °C (*see* **Notes 16** and **17**).
3. Discard the supernatant and wash 2× with high-salt wash buffer (rotate the second wash for at least 1 min in the cold room).
4. Wash 2× with PNK buffer and then resuspend in 1 ml PNK buffer (samples can be left like this at 4 °C until you are ready to proceed to the next step).

3.7 RNA 3′ End Dephosphorylation (*See* **Note 18**)

1. Discard supernatant. Resuspend the beads in 20 μl of the following mixture:

   4 μl 5× PNK pH 6.5 buffer, 0.5 μl PNK (with 3′ phosphatase activity, NEB), 0.5 μl RNasin (Promega), 15 μl water.
2. Incubate for 20 min at 37 °C in a thermomixer at 1100 rpm.
3. Wash 1× with PNK buffer.
4. Wash 1× with high-salt wash buffer (rotate wash for at least 1 min in cold room).
5. Wash 2× with PNK buffer.

3.8 L3 Linker Ligation

1. Carefully remove the supernatant and resuspend the beads in 20 μl of the following mix:

   8 μl water, 5 μl 4× ligation buffer, 1 μl T4 RNA ligase (NEB), 0.5 μl RNasin, 1.5 μl pre-adenylated linker L3-App (20 μM), 4 μl PEG400 (Sigma).
2. Incubate overnight at 16 °C in a thermomixer at 1100 rpm.

3. Add 500 μl PNK buffer.
4. Wash 2× with 1 ml high-salt wash buffer, rotating the wash for 5 min in the cold room.
5. Wash 2× with 1 ml PNK buffer. Transfer to new microtubes after the first wash and leave in 1 ml of the second wash.

3.9 5′ End Labeling

1. Collect 200 μl (20 %) of beads from the previous step (Subheading 3.8) and remove the supernatant.
2. Add 4 μl of hot PNK mix:
   0.2 μl PNK, 0.4 μl ^{32}P-γ-ATP (Perkin Elmer), 0.4 μl 10× PNK buffer, 3 μl water.
3. Incubate for 5 min at 37 °C in a thermomixer at 1100 rpm.
4. Remove the supernatant and add 20 μl of 1× NuPAGE loading buffer prepared by mixing 4× stock with water.

   Optional: Use reducing agent and antioxidant to avoid potential interference of antibodies with the protein–RNA complexes during SDS-PAGE and nitrocellulose transfer.
5. Remove the supernatant from remaining cold beads. Then add radioactively labeled beads to the cold beads. Incubate at 70 °C for 5 min in a thermomixer at 1100 rpm.
6. Place on a magnet to precipitate the beads, transfer the supernatant to a new microtube, place it again on a magnet and load the supernatant on the gel.

3.10 SDS-PAGE and Nitrocellulose Transfer

1. Load the samples on a 4–12 % NuPAGE Bis-Tris gel according to the manufacturer's instructions (*see* **Note 19**). Use 0.5 l 1× NuPAGE MOPS SDS running buffer. Also load 5 μl of a prestained protein size marker.
2. Run the gel for 50 min at 180 V.
3. Remove the dye front and discard it as solid radioactive waste (it contains the radioactive ATP).
4. Transfer the protein–RNA complexes from the gel to a Protan BA85 nitrocellulose membrane using the Western Blot wet transfer apparatus according to the manufacturer's instructions (transfer for 1 h at 30 V; do not forget to add 10 % methanol to the transfer buffer) (*see* **Note 20**).
5. After the transfer, rinse the membrane in PBS buffer, then wrap it in Saran Wrap and expose it to a film at 4 °C for ~1 h (place a fluorescent sticker next to the membrane to later align the film and the membrane).
6. Visualize the film on a phosphorimager (*see* **Note 21**).

3.11 RNA Isolation

1. Use the high-RNase condition to examine the specificity of the protein–RNA complexes (*see* **Note 22**).
2. Isolate the protein–RNA complexes from the low-RNase experiment using your autoradiograph as a mask for cutting

the respective region out of the nitrocellulose membrane. Place the membrane fragments into 1.5 ml microtubes. If a piece of membrane is too large to fit down to the bottom of the tube, cut it into several pieces before placing it into the microtube.

Optional: Re-expose the membrane after excising the bands to confirm accuracy of cutting.

3. Add 10 µl proteinase K (Roche) in 200 µl PK buffer to the nitrocellulose pieces (all should be submerged). Incubate in a thermomixer at 1100 rpm for 20 min at 37 °C.
4. Add 200 µl of PK buffer/7 M urea and incubate for further 20 min at 37 °C and 1100 rpm.
5. Collect the solution and add it together with 400 µl phenol–chloroform to a 2 ml Phase Lock Gel Heavy tube (5 PRIME) (*see* **Note 23**).
6. Incubate for 5 min at 30 °C shaking at 1100 rpm (do not vortex). Separate the phases by spinning for 5 min at 16,000 × *g* at room temperature.
7. Transfer the aqueous layer into a new microtube (be careful not to touch the gel matrix with the pipette).

   Optional: Spin again for 1 min and transfer into a new microtube.
8. Precipitate by addition of 0.75 µl GlycoBlue (Life Technologies) and 40 µl 3 M sodium acetate pH 5.5. Then mix and add 1 ml 100 % ethanol, mix again and place overnight at –20 °C (*see* **Note 24**).
9. Spin for 20 min at 21,000 × *g* at 4 °C. Remove the supernatant, wash the pellet with 0.9 ml 80 % ethanol and spin again for 5 min. Resuspend the pellet in 5 µl water and transfer to a PCR tube (*see* **Note 25**).

3.12 Reverse Transcription

1. Add the following reagents to the resuspended pellet (*see* **Notes 26** and **27**):
   1 µl Rt#clip (0.5 pmol/µl), 1 µl dNTP mix (10 mM).
2. Denature by running the following thermal program:
   70 °C for 5 min, then hold at 25 °C until the RT mix is added (see below).
3. Add 13 µl RT mix per tube:
   7 µl H_2O, 4 µl 5× RT buffer (Life Technologies), 1 µl 0.1 M DTT, 0.5 µl RNasin, 0.5 µl Superscript III (Life Technologies).
   Mix by pipetting.
   Perform the reverse transcription with the following thermal program:
   25 °C for 5 min, 42 °C for 20 min, 50 °C for 40 min, 80 °C for 5 min, then hold at 4 °C.
4. Add 1.65 µl 1 M NaOH and incubate at 98 °C for 20 min.

5. Add 20 µl 1 M HEPES pH 7.3 (*see* **Note 28**).
6. Add 350 µl TE buffer, 0.75 µl GlycoBlue and 40 µl 3 M sodium acetate pH 5.5. Mix, then add 1 ml 100 % ethanol. Mix again and precipitate overnight at −20 °C.

3.13 Gel Purification

1. Spin down for 15 min at 21,000 × *g* at 4 °C. Remove the supernatant and wash the pellet with 0.5 ml 80 % ethanol. Spin down again, remove the supernatant and resuspend the pellet in 6 µl water.
2. Add 6 µl 2× TBE-urea loading buffer to the cDNA. It is recommended, at least in initial experiments, to add loading buffer also to 6 µl DNA size marker (dilution 1/30). Heat samples to 80 °C for 5 min directly before loading.
3. Prepare 0.8 l 1× TBE running buffer and fill the upper chamber with 0.2 l and the lower chamber with 0.6 l. Use a p1000 tip to flush precipitated urea out of the wells before loading 12 µl of each sample. Leave one lane free after each sample to facilitate cutting. Load the marker into the last lane.
4. Run the 6 % TBE-urea gel (Life Technologies) for 40 min at 180 V until the lower (dark blue) dye is close to the bottom.
5. Cut off the last lane containing the size marker and stain it by incubation for 10 min shaking in 10 ml TBE buffer with 2 µl SYBR Green II (Life Technologies) stock. Wash 1× with TBE and visualize by UV transillumination. Produce a mask to guide band excision by printing the marker image scaled to 100 % (Fig. 3).
6. Together with the L3-App linker, the primer sequence accounts for 52 nt of the cDNA. The upper (light blue) dye runs at ±110–130 nt, and the lowest rim of the plastic gel cassette is at ±50 nt; these marks can be used to guide excision together with the size marker. Cut three bands at 70–80 nt, 80–150 nt and 150–200 nt. Use the schematic given in Fig. 3 as template where to cut the bands. Place each gel piece into a 1.5 ml microtube (*see* **Note 29**).
7. Add 400 µl TE and crush the gel piece into small pieces with a 1 ml syringe plunger,

   OR

   Prepare 0.5 ml microtubes by piercing a hole in the bottom using a 21G needle. Place a gel fragment inside and then place the microtubes into a 2 ml collection tube. Spin at 16,000 × *g* for 2 min.
8. Incubate in a thermomixer at 1100 rpm for 1 h at 37 °C, then place on dry ice for 2 min, and place back at 1100 rpm for 1 h at 37 °C. Transfer the liquid portion of the supernatant into a Costar SpinX column (Corning Incorporated), into which you have placed two 1 cm glass pre-filters (Whatman).

9. Spin at 16,000 × *g* for 1 min at room temperature. Collect the solution and add it together with 400 µl phenol–chloroform into a 2 ml Phase Lock Gel Heavy tube.
10. Incubate for 5 min at 30 °C shaking at 1100 rpm (do not vortex). Separate the phases by spinning for 5 min at 16,000 × *g* at room temperature.
11. Transfer the aqueous phase into a new microtube (be careful not to touch the gel matrix with the pipette). Spin again for 1 min and transfer into a new microtube.
12. Add 1 µl GlycoBlue and 40 µl 3 M sodium acetate, pH 5.5. Mix, then add 1 ml 100 % ethanol. Mix again and precipitate overnight at −20 °C.

3.14 Circularization

1. Spin down and wash with 80 % ethanol as described above and resuspend in 8 µl ligation mix:

   6.5 µl water, 0.8 µl 10× CircLigase Buffer II (Epicentre), 0.4 µl 50 mM $MnCl_2$, 0.3 µl CircLigase II (Epicentre). Transfer to PCR tubes and incubate for 1 h at 60 °C.
2. Add 30 µl oligo annealing mix:

   26 µl H_2O, 3 µl FastDigest Buffer (Fermentas), 1 µl 10 µM Cut_oligo.
3. Anneal the oligonucleotide with the following thermal program:

   95 °C for 2 min; then successive cycles of 20 s, starting from 95 °C and decreasing the temperature by 1 °C each cycle down to 25 °C; then hold at 25 °C.
4. Add 2 µl FastDigest BamHI (Fermentas) and incubate for 30 min at 37 °C, then incubate at 80 °C for 5 min.
5. Add 350 µl TE, 0.75 µl GlycoBlue, 40 µl 3 M sodium acetate, pH 5.5, and mix. Then add 1 ml 100 % ethanol. Mix again and precipitate overnight at −20 °C.

3.15 PCR Amplification

Spin down and wash the cDNA with 80 % ethanol as described above, then resuspend it in 21 µl water.

3.15.1 Optimize PCR Amplification (*See* Note 30)

1. Prepare the following PCR mix:

   1 µl cDNA, 0.25 µl primer mix of P5Solexa and P3Solexa (10 µM each), 5 µl Accuprime Supermix 1 enzyme (Life Technologies), 3.75 µl H_2O.
2. Run the following PCR:

   94 °C for 2 min, followed by *as less as possible* cycles of: (*see* **Notes 31** and **32**)

   94 °C for 15 s, 65 °C for 30 s, 68 °C for 30 s, 68 °C for 3 min, then hold at 25 °C.

3. Run 2 μl of the amplified library on capillary gel electrophoresis using the High Sensitivity D1000 Kit (Agilent Technologies) on a TapeStation system (Agilent Technologies) (*see* **Notes 33** and **34**).

3.15.2 Preparative PCR

1. From the results of the capillary gel electrophoresis, estimate the minimum number of PCR cycles to amplify 1/2 of the library (*see* **Note 35**).
2. Prepare the following PCR mix:

   10 μl cDNA (from Subheading 3.15), 9 μl H_2O, 1 μl primer mix of P5Solexa and P3Solexa (10 μM each), 20 μl Accuprime Supermix 1 enzyme.
3. If everything is fine also amplify the second 1/2 of the library with the same cycle number and combine the two PCR reactions.

3.16 Post-PCR Processing

1. Perform PCR clean-up protocol using the MinElute PCR Purification Kit (Qiagen) (*see* **Note 36**). Elute the library in 15 μl water. Measure the concentration of eluted library with NanoDrop.
2. Take ca. 500 ng (according measured concentration from the previous step) of library and dilute them in water to a total volume of 10 μl.
3. Perform size selection with the DNA 500 LabChip Kit (Perkin Elmer) on the LabChip system (*see* **Note 37**). Follow the manufacturer's guideline to perform the size selection. Cut the band in the range of 150–225 bp (the insert accounts for 20–95 bp).
4. Measure the concentration of size-selected library using the Qubit dsDNA HS Kit (Life Technologies). Use 2 μl of the sample for Qubit measurement.
5. Run 2 μl of size-selected library on capillary gel electrophoresis using the High Sensitivity D1000 Kit (Agilent Technologies) in a TapeStation system (Agilent Technologies).
6. Use the peak size from the capillary gel electrophoresis together with the Qubit measurement to estimate the molar concentration of the library (*see* **Note 38**).
7. Submit the library for high-throughput sequencing or store at −80 °C.

4 Notes

1. If you are working with a tissue with high RNase A activity, adding 1/1000 volume of ANTI-RNase (Ambion) will control the RNase conditions, without affecting the activity of RNase I.

2. The quality of the Rt#clip primers depends strongly on the individual round of synthesis (Fig. 2c). Therefore it is important to compare all primers on the same input RNA, when reordering primers or when ordering new primers for the first time.
3. The length of cross-linking should be optimized for each protein, as each RNA-binding domain cross-links with different efficiency depending on its content of aromatic amino acids and the nucleotide composition of the binding site. Try 100, 200, and 400 mJ/cm^2, then use the shortest condition that gives >70 % of the maximum signal.
4. Unless specified differently, all washes throughout the protocol are performed in a volume of 900 µl.
5. You can prepare the beads with antibody in advance and store them on ice if preparation of the cell extract takes more than 60 min.
6. For rabbit antibodies, protein A Dynabeads (Life Technologies) can work better in some cases.
7. The required amount of antibody depends on its quality and purity. This should be optimized in preliminary experiments.
8. Cell culture pellets have ±20 mg. Weighing pellets before freezing can help estimate the required volume of lysis buffer. An even better way is to determine RNA/protein concentration with NanoDrop/Bradford assay and normalize concentrations to the lowest sample. Then take 1 ml from each sample and proceed to the next step. Comparable RNA/protein concentrations should lead to more reproducible RNase digestions.
9. Sonication helps when using cell culture, as undigested viscous DNA can sometimes cause problems with the IP. It can also alleviate problems caused by mild lysis buffers or hard-to-lyse tissues. When the protein of interest is shown to have interaction with DNA, this step will be critical to reduce the protein loss. However, in most of the cases, the sonication can be excluded from the protocol.
10. Optionally, the lysate can be pre-cleared with protein A sepharose (this does not hurt, but usually makes little difference; it may reduce background when using protein A Dynabeads with certain extracts). Prepare a 30 % protein A sepharose slurry in water. Add 100 µl protein A sepharose slurry to 1.5 ml lysate and rotate for 10 min in the cold room before spinning.
11. The optimal dilution factor for the low-RNase condition depends on the batch of RNase, so in the first experiment several dilutions should be tested (Fig. 2b). It is advisable to use two low-RNase concentrations that are close to the optimized range (this could be 1:100 and 1:1000, or any other combina-

tion that works). Since there is variability between experiments, this will ensure that at least one of the concentrations will be in the right range.

12. The high-RNase control is important to monitor the specificity of the IP. Other recommended controls include a control where the RNA-binding protein is absent from the original material (such as a knockout animal or knockdown cells), a control where cross-linking is omitted and a control where no antibody is used during IP.
13. Unlike other RNases, RNase I has little base preference, and therefore cleaves after all four nucleotides. Under high-RNase conditions, the size of the radioactive band in the SDS-PAGE has to change in comparison to low-RNase conditions, confirming that the band corresponds to a protein–RNA complex. Furthermore, this experiment helps to determine the size of the immunoprecipitated RBP, as the protein will be bound to short RNAs and thus will migrate as a less diffuse band ~5 kDa above the expected molecular weight.
14. It is important to digest for exactly 3 min. Use 1.5 ml microtubes in a thermomixer for 1.5 ml microtubes to make the warming to 37 °C efficient and reproducible.
15. To test a new antibody, collect 15 μl of lysate at this step for Western blot comparison of lysate before and after IP (to visualize depletion of the protein from the lysate).
16. If monitoring depletion efficiency, place on magnet and save 15 μl supernatant for Western blot analysis (*see* **Note 15**).
17. The rotating/incubation time should always be the same to be comparable between experiments.
18. Subheadings 3.7 and 3.8 do not need to be carried out on no-UV and high-RNase controls. However, they need to be carried out on the no-antibody control to use it as a background estimate for the complete library preparation.
19. The NuPAGE gels are critical. A pour-your-own SDS-PAGE (Laemmli) changes its pH during the run, which can get to ~9.5 leading to alkaline hydrolysis of the RNA. The NuPAGE buffer system is close to pH 7. We use the NuPAGE MOPS SDS running buffer.
20. The pure nitrocellulose membrane is a little fragile, but it works better for the RNA/protein extraction step.
21. Use the shortest exposure that gives visible bands, visualize using the Image J software, and determine if the shift of the protein–RNA complex is reproducible in parallel experiments. If the signal in shortest exposure is strong enough, then expose for 15 min to a phosphorimager cassette and quantify with a Typhoon Phosphorimager (GE Healthcare) instead of Image J.

22. When performing iCLIP for the first time, use questions a) to c) to check that a specific UV cross-link and IP have been performed. Then follow points d) and e) as a guideline to properly cut the membrane and extract the complexes :
    (a) Is there a radioactive band ~5 kDa above the molecular weight of the protein in the high-RNase experiment?
    (b) Does the band disappear in the control experiments? These might include: no UV cross-link, pull down without antibody (beads only or pre-immune serum), samples from a knockout organism or knockdown cells, or an appropriate control for overexpressed tagged proteins.
    (c) Does the band move up and become more diffuse under low-RNase conditions? Because the RNA digestion is random, the RNA sizes vary more in the low RNase condition and thus the RNA–protein complexes are more heterogeneous in size. On this basis, if you are convinced of the veracity of your results, proceed to RNA isolation and amplification. Note the following guidelines:
    (d) The average molecular weight of 70 nt RNA is ~20 kDa. As the tags contain a linker of 21 nt (L3-App), the ideal position of RNA–protein complexes that will generate iCLIP cDNAs of sufficient length is ~20–60 kDa above the expected molecular weight of the protein.
    (e) The width of the excised band depends on potential other RNA–protein complexes present in the vicinity as seen in the high-RNase experiment. If none are apparent, cut a wide band of ~20–60 kDa above the molecular weight of the protein. If, however, contaminant bands are present above the size of the protein, cut only up to the size of those bands. If the contaminating bands run below your RNA–protein complex, you might consider cutting an additional band between the contaminating band and your protein–RNA complex. The RNA sequences cloned from this band can later be used to compare with those purified with the protein-of-interest.
23. Over 90 % of the radioactive signal should be removed after proteinase K treatment. This can be monitored by a Geiger counter measurement of the membrane pieces before adding proteinase K and after removing it.
24. GlycoBlue is necessary to efficiently precipitate the small quantity of RNA.
25. Remove the wash first with a p1000 and then with a p20 or p10. Try not to disturb the pellet, but if you do, spin it down again. Leave on the bench for 3 min, but no longer, with the cap open to dry. When resuspending, make sure to pipette along the back area of the tube.

26. Use only the RT primers that have worked well in test experiments. You can find a list of nine RT primers that worked well in our experiments in Subheading 2.2.
27. Do not forget a negative control. This can either be a reaction where no RNA was added to the mix, but preferably a control sample that was isolated from a piece of nitrocellulose that did not contain the protein–RNA complex (for example the no-antibody control). Use distinct primers (Rt1clip to Rt9clip) for the control and the different replicates or experiments. The different primers contain individual 4-nt barcode sequences that allow multiplexing of samples during high-throughput sequencing and control for cross-contamination between samples.
28. This will eliminate radioactivity from strongly labeled samples after the next step, and prevent RNA from interfering with subsequent reactions. It is possible to mix up to three samples that shall be multiplexed at this point. Alternatively, cDNA libraries of each sample can be amplified separately, and mixed after the PCR.
29. The 70–80 nt band is prone to producing primer artifacts in the PCR, and even if specific cDNAs are isolated, sequences are often too short to be mapped to the genome. Therefore, if binding to short RNAs (such as miRNAs) is not of interest, it is not necessary to isolate this band.
30. The test PCR in Subheading 3.15.1 is optional. If you previously prepared libraries with the same protein and you had good radioactive RNA signal, you can estimate the number of required cycles and move directly to the preparative PCR in Subheading 3.15.2.
31. Usually 15–18 cycles would be optimal. This depends on the cross-linking efficiency of the protein and can also be less or more.
32. All post-PCR work must be carried out on a specially designated bench. This cDNA must never be taken to an area where work with iCLIP RNA is done.
33. Repeat this step until samples are seen without overamplification. The least possible PCR cycles would be preferred to reduce the PCR artifacts.
34. Although a conventional gel can be used to check the amplified library, high sensitivity systems such as TapeStation and Bioanalyzer (both Agilent Technologies) are preferred.
35. Consider that you will now be amplifying 2.5 times more concentrated cDNA, therefore one cycle less is needed than in the preliminary PCR.
36. The MinElute PCR Purification Kit will remove most of residual primers. This kit is especially helpful to concentrate low-amount libraries, which in turn allows to use less PCR cycles in the amplification step.

37. The post-PCR size selection is optional, but strongly recommended to remove residual primers and unspecific products. Removing residual primers will result in higher cluster density during high-throughput sequencing. The post-PCR size selection can also be performed using a TBE urea gel, when LabChip system is not available.
38. The kits that are used in the post-PCR processing steps for selection, visualization, and quantification are optional and can be replaced by similar products. Nevertheless, high-sensitivity kits would be preferred to enable high efficiency in the high-throughput sequencing performance.

Acknowledgements

This protocol is based on previous versions that were derived with a lot of feedback and discussions in the Ule lab. We would also like to thank all iCLIP users for their feedback and comments that help to continuously improve the protocol. We acknowledge Dr. Kathi Zarnack and members of the König lab for discussions and help on the manuscript.

References

1. Kaper JM (1969) Nucleic acid-protein interactions in turnip yellow mosaic virus. Science 166(3902):248–250
2. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
3. Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res 6(2):715–732
4. Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8:479–490
5. Scherrer T, Mittal N, Janga SC et al (2010) A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS One 5:e15499
6. Tsvetanova NG, Klass DM, Salzman J et al (2010) Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. PLoS One 5:e12671
7. Castello A, Fischer B, Eichelbaum K et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-Binding proteins. Cell 149:1393–1406
8. Kwon SC, Yi H, Eichelbaum K et al (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol 20: 1122–1130
9. Baltz AG, Munschauer M, Schwanhäusser B et al (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690
10. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793
11. Chelly J, Mandel JL (2001) Monogenic causes of X-linked mental retardation. Nat Rev Genet 2:669–680
12. Mili S, Steitz JA (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA 10:1692–1694
13. Ule J, Jensen K, Mele A et al (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37:376–386
14. Ule J, Jensen KB, Ruggiu M et al (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215
15. Hafner M, Landthaler M, Burger L et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
16. Licatalosi DD, Mele A, Fak JJ et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

17. Saulière J, Murigneux V, Wang Z et al (2012) CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat Struct Mol Biol 19:1124–1131
18. König J, Zarnack K, Rot G et al (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–915
19. Sugimoto Y, König J, Hussain S et al (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol 13:R67
20. Huppertz I, Attig J, D'Ambrogio A et al (2014) iCLIP: protein-RNA interactions at nucleotide resolution. Methods 65:274–287
21. Jangi M, Boutz PL, Paul P et al (2014) Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev 28:637–651
22. Zarnack K, König J, Tajnik M et al (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. Cell 152:453–466
23. Hussain S, Sajini AA, Blanco S et al (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep 4:255–261
24. Broughton JP, Pasquinelli AE (2013) Identifying Argonaute binding sites in *Caenorhabditis elegans* using iCLIP. Methods 63:119–125
25. Bosson AD, Zamudio JR, Sharp PA (2014) Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. Mol Cell 56:347–359

Chapter 12

A Pipeline for PAR-CLIP Data Analysis

Marvin Jens

Abstract

Photo-activatable ribonucleoside cross-linking and immunoprecipitation (PAR-CLIP) is a method to detect binding sites of RNA-binding proteins (RBPs) transcriptome-wide. This chapter covers the computational analysis of the high-throughput sequencing reads generated from PAR-CLIP experiments. It explains how the reads are mutated due to UV cross-linking and how to appropriately pre-process and align them to a reference sequence. Aligned reads are then aggregated into clusters which represent putative RBP-binding sites. Mapping artifacts are a source of false positives, which can be controlled by means of a mapping decoy and adaptive quality filtering of the read clusters. A step-by-step explanation of this procedure is given. All necessary tools are open source, including the scripts presented and used in this chapter.

Key words PAR-CLIP, CLIP, UV, Cross-linking, Next-generation sequencing, High-throughput, Transcriptome, RNA, Small RNA, miRNA, mRNA, RNA-binding protein, RBP, Binding site, FLEXBAR, Adapter removal, Read mapping, BWA, False-positive filtering, BWA PSSM, Consensus-binding sites

1 Introduction

Eukaryotic gene expression is regulated at many levels. After transcription, RNA needs to be properly spliced, capped, polyadenylated, exported from the nucleus and localized, translated, and eventually degraded. All these crucial aspects of gene expression are controlled by RNA-binding proteins (RBPs) and small RNAs which interact closely with cellular transcripts throughout their life in the cell. To study this universe of post-transcriptional gene regulation it is necessary to elucidate these interactions.

Ultraviolet light can induce covalent bonds (crosslinks) between RNA and bound proteins. As proteins can be enriched with very high specificity by immunoprecipitation, the combination of cross-linking with immunoprecipitation (CLIP [1, 2]) allows to study interactions of RNA-binding proteins with endogenous transcripts.

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_12,

In recent years, the CLIP technique has been successfully combined with high-throughput sequencing (HITS-CLIP, iCLIP, PAR-iCLIP, PAR-CLIP; *see* [3, 4] for recent overviews) to enable genome-wide detection of RBP interactions with the transcriptome. This advancement requires to make a cDNA library suitable for next-generation sequencing from the fragments of RNA recovered in the CLIP experiment.

The objective of the bioinformatic analysis of such PAR-CLIP sequencing data is to reliably align the sequencing reads to a reference, filter out artifacts, and provide high-confidence binding sites and target genes. This chapter leads the reader through the major steps of PAR-CLIP [5] sequencing data analysis, from read preprocessing, over read-alignment, to read-clustering, then scoring, and filtering of these clusters to arrive at a set of high-confidence consensus-binding sites.

1.1 Experimental Aspects of PAR-CLIP Shape the Data

The different CLIP-seq methods (iCLIP, HITS-CLIP, PAR-CLIP) are all complex and differ in their details. Here, it is not possible to present the protocols in detail. However, two important aspects of the PAR-CLIP protocol need to be kept in mind when analyzing and interpreting the sequencing reads.

1.2 Thionucleoside Analogue Labeling of RNA Mutates the cDNA Sequence

In comparison to the naturally used uridine (U replaces T in the RNA alphabet), 4-thiouridine (4SU) offers a dramatically enhanced efficacy to form UV-induced cross-links [5]. This photo-activatable ribonucleoside (the "PAR" in PAR-CLIP) is added to the cell culture medium (it is also possible to label cultures of *C. elegans* animals and perform *in vivo* PAR-CLIP [6] and incorporated into newly synthesized transcripts. As a rule of thumb, HEK293 cells treated with the standard protocol may roughly contain one 4SU in 40 transcript uridine residues. This level is usually tolerated without any detectable changes in gene expression [4]. 4SU has slightly different base-pairing features than normal uridine, namely the G:U "wobble" base pair is much more stable. This means that successfully crosslinked RNA, which must contain 4SU, will produce mutations during reverse transcription into cDNA, because the "U" is often read as a "C." Empirically, it has been observed that this type of mismatch is even further enhanced for protein-cross-linked 4SU residues, which after protein digestion may still carry amino acid residues, further interfering with the reverse transcriptase.

Hence, the resulting cDNA library contains mutated sequences which no longer represent the original RNA that was bound by the RBP. This is different from random sequencing errors and complicates the read alignment. However, as the characteristic mutation of T in the reference to C is also the hallmark of cross-linking, it can be used to the advantage of the analysis. Of note, 6-thioguanosine (6SG) can be used as a (slightly less efficient) alternative to 4SU. As

6SG can also base pair with a U, it induces G:A mutations in the reads' sequences analogous to the 4SU T:C mutations. Lastly, UV light in general can induce lesions in RNA [7] and thereby cause the sequencing reads to deviate from the reference, complicating read alignment even further.

1.3 Partial Digestion of the Bound Transcripts with RNase

The standard protocol [5] uses RNase-T1 to reduce the RBP-bound transcripts to short fragments amenable to sequencing library generation. This enzyme has clear preferences to cleave before or after guanosine (G) residues. It is also limited to cleaving single stranded RNA. Thus, long stretches of RNA without guanosine or largely double-stranded structure may be digested with low efficiency and thus underrepresented in the library. On the other hand, G-rich sequences may be over-digested, yielding fragments that are too short to reliably detect their transcriptomic origin with bioinformatic tools. In consequence, the degree to which some RBP-binding sites are visible in the sequencing data depends critically on the details of the partial RNA digestion. Investing some time into optimization in the wet lab may well pay off. Also, alternative choices may be considered: RNase-I has much less sequence bias and may produce more uniformly distributed fragments, potentially benefiting the statistical power of downstream analyses. If the binding sites are highly structured, a double strand-specific RNase may also help (*see* for example [8]). Finally, in addition to the visibility of a binding site, also the shape of the coverage around a binding site, which emerges from the aligned PAR-CLIP reads, is to a large extent determined by the RNase activity (*see* also [9]). Typically, it will not be smooth as for example a CHIP-seq peak, but display pronounced changes in coverage next to guanosines in the reference sequence.

2 Sequencing Data Analysis

2.1 Scripts and Test Data

All scripts and test data used in this chapter can be obtained from *github*:

```
$ git clone https://github.com/marvin-jens/clip_analysis
```

Have a look at the README file to set up your environment.

2.2 Adapter Removal

After the CLIP experiment, the sequencing library is essentially prepared as a small RNA library: single-stranded adapters are ligated directly to the RNA fragments on both 5′ and 3′ ends before reverse transcription. The length distribution of the actual inserts highly depends on the RBP and on the partial RNase digestion (see above), but it is very common that the sequencing machine will read into the 3′ adapter sequence. At this step, it is

preferable to be conservative and remove even single nucleotides from the 3′ end if they could derive from the adapter sequence to avoid false mappings. A good tool for this is the flexible adapter remover (FLEXBAR [10]), which can be downloaded from SourceForge: http://sourceforge.net/projects/flexbar/

This is how FLEXBAR can be used to remove the 3′ adapter sequence contained in the ELAVL1 4SU PAR-CLIP test data:

```
$ echo -e '>adapter\nTCGTATGCCGTCTTCTGCTTGT' > adapter.fa

$ flexbar -a adapter.fa -ae RIGHT --adapter-threshold=2 \
   -r elavl1_4su_test_data.fastq.gz \
   -t trimmed.elavl1_4su_test_data \
   -f sanger --pre-trim-phred=3 \
   --min-read-length=15 -u 2 -z GZ --threads=8
```

The command tells FLEXBAR to look for the adapter in the right end of the read, with up to two mismatches or indels per 10 nt of adapter. As a little extra, it first removes low quality bases in the 3′ end (--pre-trim-phred), keeps reads with up to two uncalled bases (-u 2), and uses both compressed input and output. Reads that are shorter than 15 nt after adapter removal are discarded. This is a reasonable lower bound when intending to map against the human genome.

Out of the 2.5 million raw reads in the test data, only ~50 % remain in the output. But those remaining reads are not only clean of adapter sequence, they also satisfy minimal quality and length constraints.

2.3 Collapsing Identical Reads

Typically a few read sequences are repeated many times in the input. This may well reflect biology and is not a bad thing per se. However, there may be situations where the library was strongly PCR-amplified to compensate for low input. This can create many copies of the same reads which would wrongly appear as a very strong signal. To guard against such artifacts one may consider to only count each *distinct* read sequence once, regardless of how many times it appears. A real, strong biological interaction should produce a larger number of different read sequences which all go back to different original RNA fragments. Only handling distinct reads also saves time during read alignment and all later steps of the analysis because identical sequences need only be aligned once. Ideally, identical read copies are collapsed into one read, with the information about the number of copies still retained in the read name, where it can be accessed later on. A perl script (provided by Sebastian Mackowiak) can do this quite efficiently (provided you have a few gigabytes of RAM):

```
$ zcat trimmed.elavl1_4su_test_data.fastq.gz | \
   ./collapse_reads.pl elavl1_4su 2> collapsing.log | \
```

```
gzip > collapsed.elavl1_4su_test_data.fastq.gz
```

The argument passed to the *collapse_reads.pl* script is simply a prefix, used for the collapsed read names. The *stderr* stream is redirected to a log-file and contains some statistics on the reads, their length and nucleotide composition. For example, from this log, you can find that 509,547 distinct sequences were contained among the 1,322,255 adapter-trimmed reads of the test data (39 %), which is well within the normal range. If this is lower than 10 % to 5 %, this could indicate high PCR amplification of the library and too low RNA input, respectively.

The collapsed FASTQ output now includes only distinct read sequences, encoding their original copy numbers in the read name (…_x<copy-number>):

```
elavl1_4su_0_x16016
CTACAGTCCGACGATC
+
:GGGDGBEDGFFFE?F
```

2.4 Read Mapping

The collapsed reads are now ready for alignment against the reference genome ("mapping"). As the read sequences are mutated by the thiolabel cross-linking, we need to allow mismatches and indels. A good choice for mapping short reads is the Burrows Wheeler Aligner (BWA) [11]. For short reads, BWA performs alignments in two steps. However, these can be combined with UNIX pipes, to avoid storage of intermediate files. We assume that a suitable mapping index is in place (see the README in your git clone of the repository with the scripts and example data):

```
$ bwa aln -t 8 -n 1 -l 100 -k 1 reference/hg19/
hg19.fa \
    collapsed.elavl1_4su_test_data.fastq.gz 2>
    bwa_aln.log | \
   bwa samse reference/hg19/hg19.fa - \
    collapsed.elavl1_4su_test_data.fastq.gz 2>
    bwa_samse.log | \
   samtools view -hbuS -F 4 - | \
   samtools sort - hg19.elavl1_4su_test_data
$ samtools index hg19.elavl1_4su_test_data.bam
```

The first chain of commands instructs bwa to find matches in the reference with up to one edit, disables "seeding" (this makes it slower, but more sensitive) and converts the output on the fly into a sorted BAM file, a compact binary representation of the alignments. Note that unaligned reads are discarded with this command (-F4 flag for samtools). For fast random access to the aligned reads, the last command builds an index (.bai file).

2.5 Cluster Building

As the read alignments derive from a CLIP experiment, they should form small clusters in the reference genome, piling up in the regions that correspond to the parts of the transcriptome that were indeed directly bound by the probed RBP. To identify these binding sites, the information scattered over multiple individual read alignments needs to be integrated. The *clip.py* script does just that for you. It collects reads that contiguously (without a gap) cover a stretch of the reference, and jointly screens them for cross-link conversions and a number of other quality metrics:

```
$ ./clip.py -S hg19 hg19.elavl1_4su_test_data.
bam --logfile clip.log \
   > elavl1_4su_test_data.gff
```

The resulting Genome Feature Format (GFF) file contains the genomic start and end coordinates of read clusters, as well as the aggregate number of detected, cross-link-induced conversion events. The script furthermore, by default, ensures that reported clusters are supported by at least two distinct read sequences of which at least one uniquely aligns and at least one has a cross-link conversion. This is probably what you were looking for and the analysis could end here.

The problem is to what extent are you willing to trust the data? Is every cluster in this list a real binding sites? In other words, what are reasonable cutoffs on the cluster quality metrics to ensure we are not looking at mapping artifacts?

2.6 Building a Mapping Decoy

It is important to realize that there will always be mapping artifacts. Due to the mutated nature of the PAR-CLIP reads, even finding the best match in the reference is not guaranteed to also yield the *true source* of the biological interaction. However, such "shadows" of real interactions, false mappings of badly treated RNA, should intuitively be of lower quality than the real hits. In order to test this hypothesis and to be able to choose informed quality cutoffs, it is helpful to have a mapping decoy. A good choice is random sequence that should have absolutely nothing mapping to it. Essentially we are turning the problem of not knowing what a true positive looks like around, and construct a monitoring tool that allows to collect false positives, without any doubt:

```
$ ./markov.py -l 500000000 hg19.2mer > rnd_hg19.fa

$ ./markov.py -H ">chrmarkov_utr3" -l 10000000
utr3.2mer > rnd_3utr.fa

$ ./markov.py -H ">chrmarkov_intron" -l 100000000
introns.2mer > rnd_introns.fa

$ ./markov.py -H ">chrmarkov_CDS" -l 10000000
utr3.2mer > rnd_cds.fa

$ cat rnd_hg19.fa rnd_3utr.fa rnd_introns.fa
rnd_cds.fa reference/hg19/hg19.fa \
```

```
> reference/hg19/hg19_w_decoy.fa
$ bwa index -a bwtsw reference/hg19/hg19_w_
decoy.fa
```

These commands generate random sequence with the same dinucleotide frequencies as the human genome, 3′UTRs, introns, and coding sequence. For convenience, the decoy sequences can be added to the human genome as an additional "chromosome." This requires to build a new mapping index (which may take some time). To make maximal use of the mapping decoy, we need to know how likely it is going to be hit by a false mapping in comparison to the genome. As a simple assumption, we may take its relative size, which is ~20 % of the human genome reference sequence (hg19.fa). A more detailed investigation (using simulated PAR-CLIP reads) shows that it is actually closer to 10 %. With this information, we can now remap the reads to the index that includes the decoy and build clusters with *clip.py* again. The commands are essentially as above, with trivial changes wherever the new mapping index is required.

2.7 Estimating the False Discovery Rate

We can now count, how many times a cluster was built from alignments to the mapping decoy. These are guaranteed to represent false positives:

```
$ grep --count markov elavl1_4su_test_data.gff
241
$ grep --count --invert-match markov elavl1_4su_
test_data.gff
3821
```

Note that the actual number may differ a bit due to the randomness in the generation of the decoy sequence. As we have found the decoy to have a 10 % probability to be hit by a false positive, compared to the genome, the number of false positives in the genome could be as high as 2410. Having 3,821 clusters reported in the genome, this indicates a very high false discovery rate (FDR, false positives divided by all positives) of more than 60 % in the otherwise unfiltered PAR-CLIP cluster set.

2.8 Reducing the False Discovery Rate by Filtering

If indeed the false positives look *worse* than the true positives in some quantifiable way, we may be able to reduce the FDR by demanding some minimal quality. A prime candidate is the number of crosslink conversion (T:C mismatch) events, which conveniently is recorded in the score column (6) of the GFF file:

```
$ cat elavl1_4su_test_data.gff | ./scorethresh.
py 6 3 | grep -c markov
33
```

```
$ cat elavl1_4su_test_data.gff | ./scorethresh.
py 6 3 | grep -vc markov

1061
```

This indicates that demanding at least three independent (the reads have been collapsed) conversion events is sufficient to half the FDR to 31 %. That is still too high for most purposes. But the list of clusters is also quickly shrinking. Apparently we are throwing away a lot of weak signal as well. There are a number of metrics that should correlate with the reliability of the mapping. Foremost, this would be the MAPQ reported by BWA. But an even simpler object is the length of a cluster. Longer clusters are either derived from longer reads (which are more reliable to align), or a number of overlapping, different reads. In both cases, long clusters should be more reliable:

```
$ cat elavl1_4su_test_data.gff | ./scorethresh.
py 6 2 | \

   ./sum.py -4,5 | ./scorethresh.py 10 21 | grep
-c markov

9

$ cat elavl1_4su_test_data.gff | ./scorethresh.
py 6 2 | \

   ./sum.py -4,5 | ./scorethresh.py 10 21 | grep
-vc markov

1881
```

By jointly demanding two or more conversion events and a length of at least 21 nt (the difference of end and start coordinates in columns 4,5) the FDR is now at ~4.8 %, below the 5 % threshold that is typically considered acceptable. Note how 1881 clusters (roughly one-third of the input) survive these filters. This cluster set should represent a much more reliable input for any downstream analysis. A priori, it is far from obvious what is the best way to filter a cluster set for quality. However, with the mapping decoy in place, the procedure can be automated. The *clip.py* script can be instructed to report all kinds of additional cluster scores to a separate file. The *fdr.py* script can then act on this output and, by the use of a mapping decoy, automatically select the optimal combination of quality criteria to filter on, in order to satisfy a given FDR limit (5 % by default):

```
   $ ./clip.py -S hg19 hg19.elavl1_4su_test_
data.bam \

   -l clip.log --fp_flag markov --cluster-stats
elavl1_4su_test_data.tsv \

   > elavl1_4su_test_data.gff
```

```
$ ./fdr.py elavl1_4su_test_data.tsv
elavl1_4su_test_data.gff \
--cutoff 0.05 --decoy 0.1 --logfile fdr.log \
> fdr_filtered.elavl1_4su_test_data.gff \
2> filtered_out.elavl1_4su_test_data.gff
```

With the test data, this procedure finds that demanding a minimum length of 24 nt is sufficient to reduce the FDR below 5 % and is able to retain 2051 clusters in the genome (you find the details in fdr.log).

Another useful property of this approach is that the criteria that were selected for filtering, and the stringency that was required to reduce the FDR, speak to the quality of the experiment: length may be selected because many reads are too short for reliable mapping, indicating possible over digestion with RNase. The fraction of reads with a crosslink conversion ("signature_density") may be selected if crosslinking was specific and efficient.

3 Further Improvements and Variations of the Analysis

3.1 Replicates

The analysis can very much benefit from using multiple biological replicates and building consensus clusters. The *clip.py* script can operate on multiple sorted bam files simultaneously and supports simple consensus calling rules:

```
$ ./clip.py rep1.bam rep2.bam rep3.bam --min_
support=2 --require_conversions
```

Will only report clusters that are supported by reads with crosslink conversions deriving from at least two out of the three provided replicates.

3.2 Mapping with a PAR-CLIP Error Model

A recent publication describes BWA-PSSM [12], a BWA-based mapper that does not align read sequences, but position specific weight matrices. It thereby allows to use a custom error model: occurrences of a "C" in the read can have a lower mismatch cost, when paired with a "T" in the reference. An initial evaluation indicates that the mapping of highly mutated PAR-CLIP reads can substantially benefit from this approach. Importantly, the uniqueness of an alignment is now represented as the (posterior) probability that the alignment is the true source of the read *given the cross-link conversion error model*. This allows a much more accurate delineation of true mappings from PAR-CLIP induced artifacts and is reflected in the observation that *fdr.py* often selects "map_pp," the posterior probability for the alignment, when BWA-PSSM is used for mapping. We successfully used it with its default PAR-CLIP error model on *C. elegans* in vivo PAR-CLIP data [8]:

```
$ bwa pssm -t 8 -m 800 -G bwa-pssm/PAR_CLIP_
error_model.txt […]
```

This runs BWA-PSSM on eight parallel threads, with high sensitivity (by keeping up to 800 entries in its queue of possible matches).

3.3 Removing Systematic False Positives

Cellular RNA contains large amounts of ribosomal and transfer RNAs, and these sequences are often observed in PAR-CLIP experiments. Depending on the RBP you are investigating, they should probably be considered contaminants and excluded from the analysis. A straightforward solution is to use the *clip.py* support for masking regions of the genome (such as rDNA clusters) with the -M command-line option.

As the cDNA library preparation follows essentially the protocol for small RNA sequencing, it is prone to capture other 5′ monophosphate-bearing RNAs such as miRNAs. Again, depending on your RBP of interest, you may consider removing reads that derive from known miRNA sequences. At the very least, it is advisable to take binding sites in miRNAs with a grain of salt and always a good idea to compare to another PAR-CLIP library to convince yourself that the signal is specific.

3.4 Other Tools

There is no single "best" way to analyze your data. It always depends on the exact question you are trying to answer and on the details of the project. There are other excellent tools available to analyze PAR-CLIP data, that may be more suited to your needs or easier to work with. Especially noteworthy are PARalyzer [13], which features a kernel density estimation-based segmentation of clusters into binding sites, and CLIPz [14], a web-browser-accessible environment for data analysis and exploration that does not require use of command-line tools.

Acknowledgments

The procedures for PAR-CLIP data analysis presented in this chapter have been developed over the past 5 years in the Rajewsky lab and have received input and work from many people. Among them are Jonas Maaskola, Sebastian Mackowiak, Minnie Zhuo Fang, Andranik Ivanov, and Nikolaus Rajewsky. They have also been adapted and developed in close collaboration with the wet lab, where PAR-CLIP experiments have been performed by Svetlana Lebedeva, Agnieszka Rybak-Wolf, Yasuhiro Murakawa (Landthaler lab), Anna-Carina Jungkamp, and Stefanie Grosswendt. The author further acknowledges many fruitful discussions with Markus Landthaler.

References

1. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215
2. Ule J, Jensen K, Mele A, Darnell RB (2005) CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37:376–386
3. McHugh CA, Russell P, Guttman M (2014) Methods for comprehensive experimental identification of RNA-protein interactions. Genome Biol 15:203
4. Milek M, Wyler E, Landthaler M (2012) Transcriptome-wide analysis of protein-RNA interactions using high-throughput sequencing. Semin Cell Dev Biol 23:206–212
5. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp A-C, Munschauer M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
6. Jungkamp A-C, Stoeckius M, Mecenas D, Grün D, Mastrobuoni G, Kempa S, Rajewsky N (2011) In vivo and transcriptome-wide identification of RNA binding protein target sites. Mol Cell 44:828–840
7. Kladwang W, Hum J, Das R (2012) Ultraviolet shadowing of RNA can cause significant chemical damage in seconds. Sci Rep 2:517
8. Rybak-Wolf A, Jens M, Murakawa Y, Herzog M, Landthaler M, Rajewsky N (2014) A variety of dicer substrates in human and C. elegans. Cell 159:1153–1167
9. Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8:559–564
10. Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. Biology (Basel) 1:895–905
11. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
12. Kerpedjiev P, Frellsen J, Lindgreen S, Krogh A (2014) Adaptable probabilistic mapping of short reads using position specific scoring matrices. BMC Bioinformatics 15:100
13. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U (2011) PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol 12:R79
14. Khorshid M, Rodak C, Zavolan M (2011) CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic Acids Res 39:D245–D252

Part IV

Non-coding RNAs

Chapter 13

Capture and Identification of miRNA Targets by Biotin Pulldown and RNA-seq

Shen Mynn Tan and Judy Lieberman

Abstract

MicroRNAs (miRNAs) are small noncoding RNAs that regulate the stability and expression of target RNAs in a sequence-dependent manner. Identifying miRNA-regulated genes is key to understanding miRNA function. Here, we describe an unbiased biochemical pulldown method to identify with high-specificity miRNA targets. Regulated transcripts are enriched in streptavidin-captured mRNAs that bind to a transfected biotinylated miRNA mimic. The method is relatively simple, does not involve cross-linking and can be performed with only a million cells. Addition of an on-bead RNase digestion step also identifies miRNA recognition elements (MRE).

Key words MicroRNA, Noncoding RNA, miRNA recognition element, Target identification

1 Introduction

More than a thousand miRNAs are expressed in human cells, each of which can regulate hundreds of coding and noncoding RNAs. miRNAs regulate as many as 90 % of mRNAs [1–3]. miRNAs recognize the genes they regulate by base pairing. Because the miRNA sequence is short (~22 nt) and is not perfectly complementary to target gene sequences, predicting regulated genes is not straightforward. Sequence properties of miRNA-regulated targets have led to a series of canonical rules that are the basis for algorithms that predict miRNA targets. These rules emphasize the importance of exact base pairing of a sequence in the 3′UTR of the regulated transcript to the seed region (nucleotides 2–9) of the miRNA. However, many miRNA-target RNA interactions do not obey the canonical rules. Thus even the best target-prediction algorithms have poor sensitivity and specificity for predicting regulated genes [4]. Since the transcripts of genes that are regulated by a miRNA are generally down-regulated by miRNA over-expression or up-regulated by miRNA knockdown [4], the predictive power of algorithms can be enhanced by combining lists of predicted

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_13,

targets with experimental data on changes in gene expression that occur with manipulating miRNA expression in cells of interest.

Biochemical isolation of miRNA-induced silencing complexes using antibodies to AGO or other protein components (miRISC) and cross-linking to fix AGO-associated RNAs with techniques such as HITS-CLIP [5, 6], CLIP-seq [7], PAR-CLIP [8], and CLASH [9] provides a global snapshot of miRNA-RNA interactions. However, these methods require large numbers of cells and are technically challenging to perform. They also tend to be less effective for identifying the targets of poorly expressed miRNAs. Identifying within these data the targets of an individual miRNA often requires assumptions about miRNA recognition elements, which introduces considerable bias. The CLASH method has the advantage that it directly identifies miRNA:target miRNA recognition element (MRE) pairs by ligating them and then sequencing the ligated gene product. A recent study, however, suggests that it may be possible to identify such pairs even without adding a ligation step, because of spontaneous ligation by endogenous ligases during the reaction in cell lysates [10]. The UV cross-linking step central to these methods is inefficient, introduces background, and is known to increase the false discovery rate for genome-wide studies of RNA-protein interactions [11].

Streptavidin beads can be used to capture RNAs bound to a transfected miRNA mimic biotinylated at the 3′-end of the active strand [12–19]. These pulldown procedures can identify the transcripts regulated by a specific miRNA in a specific cell type without bias and with a high degree of specificity (70–90 % in several studies) [12–14]. The transfected miRNA mimic is incorporated into the miRISC and is active. The pulldown procedure isolates miRNA:mRNA pairs associated with the miRISC since AGO depletion largely eliminates recovery of specific target mRNAs [13]. These protocols do not require cross-linking and can be performed with just a million transfected cells. When the abundance of RNAs bound to the biotinylated mimic of interest is compared to those bound to a control biotinylated miRNA mimic by real-time PCR, microarray or RNA-seq analysis, hundreds of potential targets are identified. When they are examined experimentally, the great majority of these are *bona fide*. Here, we detail our optimized version of a pulldown protocol that identifies coding and non-coding RNAs regulated by a miRNA. It is still unclear how much the set of miRNA-regulated genes or the extent of gene regulation of each gene varies in different cell contexts. However, transcripts that might be regulated in one cell type may not be identified by the pulldown in another cell type either because the miRNA is not as abundant, the target is not expressed, or the other expressed transcripts in the cell compete for miRNA binding [20]. We also describe a modification that enables MREs to be identified. Because the number of genes that can in principle be regulated by an individual miRNA is large, figuring out the biological functions of a

miRNA from the lists of potential target genes is difficult. Here we describe a bioinformatics strategy that has enabled us to interpret the biological meaning of pulled down genes to rapidly (with minimal trial and error) identify overall miRNA function in cells.

2 Materials

Unless otherwise indicated, all solutions are prepared with nuclease-free water and nuclease-free grade reagents (Ambion, USA). Store all reagents on ice when in use; we recommend that fresh buffers be used for each experiment, unless indicated otherwise. Diligently follow all waste disposal regulations when disposing of waste materials (*see* **Note 1**).

2.1 Transfection and miRNA Mimics

1. Transfection reagent: DharmaFECT 1 (Dharmacon, USA).
2. 3′ biotinylated miRNA mimics: Control—cel-miR-67, and experimental (e.g., hsa-miR-522), with Dharmacon's proprietary modifications. Biotin is coupled to the 3′-end of the active strand of the miRNA (Dharmacon, USA).

2.2 Bead Preparation and Pulldown

1. Streptavidin magnetic beads: Dynabeads® M-280 (Life Technologies, USA).
2. Magnetic rack: DynaMag™-2 Magnet (Life Technologies, USA).
3. Tube rotator in a cold room.
4. Wash buffer: 5 mM Tris–HCl (pH 7), 500 μM EDTA, 1 M NaCl. Add 75 μl of 1 M Tris–HCl (pH 7), 15 μl of 0.5 M EDTA and 3 ml of 5 M NaCl to 12 ml of nuclease-free water. Store at room temperature.
5. Solution A: 50 mM NaCl, 100 mM NaOH. Add 100 μl of 5 M NaCl and 100 μl of 10 N NaOH to 9.8 ml of nuclease-free water. Store at room temperature.
6. Solution B: 100 nM NaCl. Add 100 μl of 5 M NaCl to 4.9 ml of nuclease-free water. Store at room temperature.
7. Phosphate-buffered saline (PBS).
8. TrypLE (Life Technologies, USA).
9. Lysis buffer: 20 mM Tris–HCl (pH 7), 100 mM KCl, 5 mM $MgCl_2$, 25 mM EDTA, 0.3 % NP-40 (Fluka, USA) with 1× Proteinase Inhibitor cocktail. Add 1 ml of 1 M Tris–HCl (pH 7), 2.5 ml of 2 M KCl, 250 μl of 1 M $MgCl_2$, 2.5 ml of 0.5 M EDTA and 150 μl of NP-40 to 43.6 ml of nuclease-free water. Add 1 tablet of Proteinase Inhibitor cocktail (Roche, USA) and mix gently by inversion till completely dissolved.
10. Blocking buffer: 1 mg/ml Ultrapure BSA, 200 μg/ml Yeast t-RNA in lysis buffer. Add 20 μl of 50 mg/ml Ultrapure BSA

and 20 μl of 10 mg/ml Yeast t-RNA (both from Ambion, USA) to 1 ml of lysis buffer. Scale up as needed.

11. Lysis buffer plus: 2.5 mg/ml Ficoll PM400, 7.5 mg/ml Ficoll PM70, 250 μg/ml Dextran Sulfate 670k, 200 U/ml RNaseOUT™, 100 U/ml SUPERase.In™ in Lysis Buffer. Weigh 2.5 mg of dextran sulfate 670k (Fluka, USA), 25 mg of Ficoll PM400, and 75 mg of Ficoll PM70 (GE Healthcare, USA) in a 50 ml tube, add 10 ml of lysis buffer and mix gently by inversion till completely dissolved. Add 50 μl of 40 U/μl RNaseOUT™ and 50 μl of 20 U/μl SUPERase.In™ (Life Technologies, USA) and mix by inversion. Scale up as needed.
12. RNase lysis buffer: 2.5 mg/ml Ficoll PM400, 7.5 mg/ml Ficoll PM70, 250 μg/ml dextran sulfate 670k, 25,000 U/ml RNase T1 in lysis buffer. Weigh 500 μg of dextran sulfate 670k (Fluka, USA), 5 mg of Ficoll PM400, and 15 mg of Ficoll PM70 (GE Healthcare, USA) in a 50 ml tube, add 2 ml of lysis buffer and mix gently by inversion till completely dissolved. Add 50 μl of 1000 U/μl RNase T1 (Fermentas, USA) and mix by inversion. Scale up as needed.

2.3 RNA Precipitation and Cleanup

1. Trizol LS (Life Technologies, USA).
2. Phenol.
3. Chloroform.
4. GlycoBlue (Ambion, USA).
5. Ethanol.
6. Phenol:chloroform:isoamyl (25:24:1).
7. Chloroform:isoamyl (24:1).
8. 3 M sodium acetate pH 5.5 (Ambion, USA).
9. NucAway spin columns (Ambion, USA).
10. PNK buffer and enzyme (NEB, USA).

2.4 RNA-seq Library Preparation

1. Ribo-Zero rRNA removal kit (Epicentre, USA).
2. NEBNext Ion Torrent mRNA library preparation kit (NEB, USA) or equivalent for the Ion Torrent platform, for target transcript identification.
3. NEBNext Illumina small RNA library preparation kit (NEB, USA) or equivalent for the Illumina platform, for target MRE identification.

2.5 Real-Time PCR

1. High-capacity cDNA archive kit (ABI, USA).
2. Fast SYBR green mastermix (Bio-Rad, USA).

2.6 MRE Validation

1. psiCHECK2 plasmid (Promega, USA).
2. NotI and XhoI restriction enzymes and buffers (NEB, USA).

3. CIP (NEB, USA).
4. Quick ligation kit (NEB, USA).
5. One Shot TOP10 competent cells (Life Technologies, USA).
6. Phusion HF PCR kit (NEB, USA).
7. Dual-luciferase reporter assay system (Promega, USA).

3 Methods

3.1 Transfection

1. Optimize the miRNA transfection protocol for your cell line of interest. Transfect enough cells with 50 μM of biotinylated miRNA mimic, to harvest between 1 and 4 million cells for each miRNA (Control and Experimental) at 50 % confluency for lysis (*see* **Note 2**).
2. For MDA-MB-468 breast cancer cell lines, we transfect 1.5 million cells with DharmaFECT 1 and 50 μM of biotinylated miRNA mimic in a 60 mm tissue culture dish as per the manufacturer's instructions, and harvest the cells for the pull-down experiment 20 h after transfection (*see* **Note 3**).

3.2 Bead Preparation

Perform all steps at room temperature. Wash buffer, solution A, and solution B can be kept for up to a month at room temperature.

1. Resuspend magnetic beads (50 μl per 1.5 ml Eppendorf tube per experiment or sample) in 1 ml of wash buffer (maximum of 300 μl of beads per 1 ml of buffer).
2. Mix by inversion and place tubes on magnetic rack for 1 min. Remove wash buffer and repeat twice (for a total of three times).
3. Resuspend beads in solution A.
4. Mix by inversion and place tubes on magnetic rack for 1 min. Remove solution A and repeat once (for a total of two times).
5. Resuspend beads in solution B.
6. Mix by inversion and place tubes on magnetic rack for 1 min. Remove solution B completely.
7. Resuspend beads in blocking buffer (maximum of an equivalent of 150 μl of beads per 1 ml of buffer).
8. Leave tubes on rotator in 4 °C for a minimum of 2 h.

   Perform the following steps during the 20-min cell lysis incubation step in Subheading 3.3 below
9. Place tubes on magnetic rack for 1 min and remove blocking buffer.
10. Resuspend beads in lysis buffer.
11. Mix by inversion and place on magnetic rack for 1 min. Remove lysis buffer and repeat once (for a total of two times).

12. Mix by inversion and place on magnetic rack for 1 min. Remove lysis buffer completely.
13. Combine all magnetic beads and resuspend in lysis buffer plus (110 µl of lysis buffer plus for each sample).
14. For each sample, aliquot 100 µl of the beads/lysis buffer plus mixture into a fresh 1.5 ml Eppendorf tube and keep on ice.

3.3 Cell Lysis and Pulldown Incubation

1. Wash all plates with 1 ml of ice-cold PBS each, and remove PBS completely.
2. Add 500 µl of TrypLE per plate and place in the incubator for 5–15 min (depending on cell line), until cells detach completely from the tissue culture surface. Pipette up and down as necessary.
3. Neutralize TrypLE by adding 1 ml of medium with serum and transfer to a 1.5 ml Eppendorf tube.
4. Spin tubes at 4 °C at 500 × *g* for 5 min, discard supernatant, and keep cell pellet.
5. Resuspend cell pellet in 1 ml of ice-cold PBS.
6. Spin tubes at 4 °C at 500 × *g* for 5 min, remove PBS, and keep cell pellet. Repeat PBS wash once (for a total of two times), taking care not to touch the cell pellet.
7. Remove PBS completely and resuspend cell pellet in 500 µl of ice-cold lysis buffer plus (*see* **Note 4**).
8. Incubate on ice for 20 min, inverting occasionally. Wash beads that are in blocking buffer during this step (*see* Subheading 3.2).
9. Spin tubes at 4 °C at 5000 × *g* for 5 min to pellet nuclei.
10. Remove post-nuclear supernatant, taking care not to touch the nuclear pellet. For each sample, set aside 50 µl of supernatant (cytoplasmic lysate) at 4 °C for analysis of total cellular RNA.
11. Add 400 µl of supernatant (cytoplasmic lysate) to 100 µl of prepared beads in lysis buffer plus for each tube (as prepared in Subheading 3.2).
12. Make sure that all tubes are sealed, and leave on rotator in 4 °C for 4 h.

3.4 Bead Washing and RNA Precipitation

1. Place bead-containing tubes on magnetic rack for 1 min.
2. Remove cytoplasmic lysate from all tubes and add 1 ml of lysis buffer.
3. Mix by inversion and place on magnetic rack for 1 min. Remove lysis buffer and repeat wash steps four more times (for a total of five times).
4. For the last wash step, mix by inversion and place on magnetic rack for 2 min. Remove lysis buffer completely.

5. For each bead-containing sample, resuspend beads in 100 µl of lysis buffer and transfer to a fresh 1.5 ml eppendorf tube. For the total RNA sample, add 50 µl lysis buffer (to achieve a final volume of 100 µl), and treat these samples like the bead-containing samples in subsequent steps.
6. Add 500 µl of TRIzol LS to each tube and mix vigorously.
7. Incubate at room temperature for 5 min, mixing occasionally.
8. Add 100 µl of chloroform to each tube and mix vigorously.
9. Spin tubes at 4 °C at >16,000 × *g* for 15 min.
10. For each sample, transfer the aqueous phase (around 250 µl) to a fresh 1.5 ml Eppendorf tube, taking care not to touch the interphase.
11. For each tube, add 5 µl of GlycoBlue, and subsequently 850 µl of 100 % ethanol.
12. Mix vigorously and precipitate overnight at −20 °C.

3.5 RNA Cleanup

1. The next day, spin tubes at 4 °C at >16,000 × *g* for 30 min. The blue pellet contains the pulldown RNA. Remove ethanol without touching the pellet.
2. For each sample, add 1 ml of 75 % ethanol and mix by inversion.
3. Spin tubes at 4 °C at >16,000 × *g* for 15 min. Remove ethanol without touching the pellet and repeat 75 % ethanol wash step once (for a total of two times).
4. Remove ethanol and dry blue RNA pellets in an open fume hood, taking care not to over-dry them as this interferes with resuspension.
5. Resuspend RNA pellets in 200 µl of nuclease-free water and place on ice.
6. For each tube, add 200 µl of phenol:chloroform:isoamyl (25:24:1).
7. Mix vigorously and spin at >16,000 × *g* for 5 min at room temperature.
8. Transfer the aqueous phase (around 200 µl) to a fresh 1.5 ml Eppendorf tube, taking care not to touch the interphase.
9. For each tube, add 200 µl of chloroform:isoamyl (24:1).
10. Mix vigorously and spin at >16,000 × *g* for 5 min at room temperature.
11. Transfer the aqueous phase (around 200 µl) to a fresh 1.5 ml eppendorf tube, taking care not to touch the interphase.
12. For each tube, add 20 µl of 3 M sodium acetate pH 5.5, 4 µl of GlycoBlue, and subsequently 1 ml of 100 % ethanol.
13. Mix vigorously and precipitate overnight at −80 °C.

14. The next day, spin tubes at 4 °C at >16,000 × *g* for 30 min. The blue pellet contains the RNA. Remove ethanol without touching the pellet.
15. For each sample, add 1 ml of 75 % ethanol and mix by inversion.
16. Spin tubes at 4 °C at >16,000 × *g* for 15 min. Remove ethanol without touching the pellet and repeat 75 % ethanol wash step once (for a total of two times).
17. Remove ethanol and dry blue RNA pellets in open fume hood, taking care not to over-dry them as this interferes with resuspension.
18. Resuspend RNA pellets in 50 μl of nuclease-free water and place on ice. The pulldown RNA is now ready for downstream analyses (*see* **Note 5**).

3.6 Pulldown Analysis

1. The pulled down RNA can be analyzed by microarray or RNA-seq. This chapter provides the details for RNA-seq (for a microarray protocol, *see* [12]). RNA-seq has the advantage that it provides information for bound RNAs of all classes including ncRNAs without being limited to particular splice variants of mRNAs probed on the array. It can also distinguish genes from highly homologous pseudogenes and alternately spliced or polyadenylated transcripts. However, library preparation is more complicated, since very abundant rRNAs need to be depleted or they will dominate the reads. Moreover the analysis of RNA-seq data is more difficult than microarray data. Essentially for either method, the fold enrichment of the miRNA targets is calculated as the normalized ratio of the experimental pulldown signal to control pulldown signal (*see* **Note 6**). For microarray data, normalization of the ratio of the experimental vs. control pulldown to the ratio of the experimental vs control input samples greatly improved the identification of bona fide targets [12]. This improvement may have been for two reasons. First, background binding of mRNAs is enhanced when they are more abundant and normalization to the input takes this bias into account. Second, transfection of the miRNA mimic reduces the amount of target mRNA. Therefore, for most targets, whose mRNA is reduced by transfection of the miRNA, the normalized enrichment ratio (experimental PD/control PD)/(experimental input/control input) goes up because the numerator increases, while the denominator decreases. Biological replicates or triplicates are required to achieve statistically meaningful results. The choice of a cutoff to separate targets from other genes is in some ways arbitrary. We use a fold enrichment cutoff of 2 and p-value <0.01, which results in a target gene list of a few hundred RNAs (*see* **Note 7**). To identify miRNA targets using microarrays, we first normalize the data (platform dependent; for

Illumina Beadchips, we use cubic spline without background normalization), and use a differential gene expression approach that compares the normalized signal of control and experimental pulldowns for each gene to calculate fold enrichment and *p*-values. This can be performed using the NIA Array Analysis software available online (http://lgsun.grc.nia.nih.gov/ANOVA/index.html) [13].

2. For RNA-seq, the coverage required for miRNA target identification can be lower than transcriptome-wide differential expression RNA-seq analyses, if we assume that the number of targets for a miRNA that we can detect is in the thousands at most (i.e., at most a few percent of the transcriptome). We recommend at least biological replicates, with more than 200 million uniquely-mapped bases per replicate. This should be achievable with most RNA-seq platforms, for example Ion Torrent from Life Technologies, or even with multiplexing using higher throughput equipment (for example the HiSeq platforms from Illumina).

3.7 Pulldown-seq Library Prep and Analysis

For Pulldown-seq library preparations, the use of straightforward kits and rRNA removal are strongly recommended. For Pulldown-seq on the Ion Torrent platform, we performed Ribo-Zero rRNA removal before preparing the library with the NEBNext mRNA library preparation kit with Ion Torrent-specific primers, following the manufacturer's protocol.

The bioinformatics analysis workflow for Pulldown-seq on the Ion Torrent platform as previously published is as follows [13]:

1. Screen pre-alignment and post-alignment libraries for quality, specificity of mapping and containment sequences using FASTQC, RSeQC and RNA-SeQC.
2. Trim low quality bases, homopolymer sequences and sequences matching the first 13 bases, and the reverse complement of the adapter sequences for Ion Torrent using cutadapt, and discard trimmed reads smaller than 30 nt.
3. Align trimmed reads to the latest build of the human genome using Novoalign (with the parameters: -H -k -n 250 -F STDFQ -r all 10 -e 10 -g 15 -x 4), keeping only uniquely mapped reads.
4. Generate post-alignment gene counts using htseq-count, with the counts aggregated by gene_id.
5. Call differentially expressed (or in this case differentially enriched) transcripts using the default options of DESeq, with the fold enrichment and *p*-value cutoffs at 2 and 0.05, respectively.

3.8 Target Validation

1. The first step to validate miRNA targets is to verify the microarray or RNA-seq enrichment in the pulldown by real-time PCR quantification of the RNA in the miRNA and control pulldown relative to its level in total cellular RNA, normalized to a house-

keeping gene mRNA such as *GAPDH*. To do this we synthesize cDNA from 2 μl of pulldown RNA from both control and experimental miRNA pulldowns in a total reaction volume of 30 μl, and perform SYBR green real-time PCR on tenfold diluted cDNA with custom-designed primers for a set of 10 or more target genes. These are selected from the microarray or RNA-seq results to represent the spectrum of fold enrichment and *p*-values. The real-time PCR fold enrichment values are then compared to the microarray or RNA-seq results for each specific target to obtain the R-squared value (*see* **Notes 5**, **8**, and **9**).

2. Next we determine whether transfection of the miRNA decreases the target gene mRNA and protein. For mRNA expression, we perform real-time PCR with cDNA synthesized from total RNA from both control and experimental miRNA transfected cells and analyze the reduction of expression for each specific target (*see* **Note 10**). The proportion of all the pulldown-predicted miRNA target RNAs that are downregulated can also be analyzed using microarrays [12–14] or RNA-seq. For protein expression, western blots of cell extracts, comparing cells transfected with control and experimental miRNAs, are probed with antibodies to proteins of interest [12–14]. Because western blots are not always quantitative, it is a good idea to compare protein levels using serial dilutions of samples, with a range of cell equivalents in each lane. The optimal time to examine target gene expression may be later than what was optimized for the pulldown, since more time is needed to downregulate gene expression. For most cells and miRNAs, testing gene regulation should be performed 48–72 h after transfection, with the later time point preferred for some genes.

3.9 RNase Treatment for Identification of MREs

The addition of RNase treatment to the above pulldown protocol will enable the identification of MREs via RNA-seq, termed IMPACT-seq for *i*dentification of *M*REs by *p*ulldown and *a*lignment of *c*aptive *t*ranscript-*seq*uencing [13]. If this is desired, please perform all steps from Subheadings 3.1–3.6, but replace Subheading 3.5 with this section. As this protocol introduces additional wash steps and removal of RNAs below 20 nt with NucAway spin columns, we generally combine the RNA for three samples for each experiment per RNA-seq reaction.

1. Remove cytoplasmic lysate from all tubes and add 1 ml of lysis buffer.
2. Mix by inversion and place on magnetic rack for 1 min. Remove lysis buffer and repeat wash steps four more times (for a total of five times). Remove lysis buffer for the final wash completely.
3. For each bead-containing sample, resuspend beads in 400 μl of ice-cold RNase lysis buffer, transfer to a fresh 1.5 ml Eppendorf tube and place on ice.

4. When all tubes are ready, incubate them in a 37 °C water bath or heat block for 10 min with occasional inversions to ensure an even temperature in each tube.
5. Place all tubes on ice for 5 min.
6. Place tubes on magnetic rack for 1 min, and remove RNase lysis buffer completely.
7. Add 1 ml of lysis buffer for each tube, mix by inversion and place on magnetic rack for 1 min. Remove lysis buffer and repeat wash steps four more times (for a total of five times).
8. For the last wash step, mix by inversion and place on magnetic rack for 2 min. Remove lysis buffer completely.
9. For each set of bead-containing samples, resuspend combined beads (usually three tubes) in 100 μl of lysis buffer and transfer to a fresh 1.5 ml Eppendorf tube.
10. Add 500 μl of TRIzol LS to each tube and mix vigorously.
11. Incubate at room temperature for 5 min, mixing occasionally.
12. Add 100 μl of chloroform to each tube and mix vigorously.
13. Spin tubes at 4 °C at >16,000 × *g* for 15 min.
14. For each sample, transfer the aqueous phase (around 250 μl) to a fresh 1.5 ml Eppendorf tube, taking care not to touch the interphase.
15. For each tube, add 5 μl of GlycoBlue, and subsequently 850 μl of 100 % ethanol.
16. Mix vigorously and precipitate overnight at −20 °C.
17. The next day, spin tubes at 4 °C at >16,000 × *g* for 30 min. The blue pellet contains the MRE RNA. Remove ethanol without touching the pellet.
18. For each sample, add 1 ml of 75 % ethanol and mix by inversion.
19. Spin tubes at 4 °C at >16,000 × *g* for 15 min. Remove ethanol without touching the pellet and repeat 75 % ethanol wash step once (for a total of two times).
20. Remove ethanol and dry blue RNA pellets in an open fume hood, taking care not to over-dry them as this interferes with resuspension.
21. Resuspend RNA pellets in 50 μl of nuclease-free water and place on ice.
22. Tap the NucAway columns to settle dry gel onto the bottom of the spin column.
23. Hydrate the columns with 650 μl of nuclease-free water, cap, vortex, and tap out the air bubbles. Leave the columns at room temperature for 10 min.
24. Place each column in a 2 ml collection tube and centrifuge the column at 700 × *g* for 2 min.

25. Discard the collection tubes, and place each column in a fresh 1.5 ml Eppendorf tube.
26. Add 50 μl of RNA for each sample onto the hydrated gel in each column slowly, and spin immediately at 700 × *g* for 2 min.
27. Discard the columns and top up size-selected RNA to 174 μl with nuclease-free water.
28. To each tube of 174 μl of size-selected RNA, add 20 μl of PNK buffer and 4 μl of PNK enzyme, as well as 1 μl each of RNase OUT Superase IN.
29. Incubate at 37 °C for 1 h.
30. Add 200 μl of phenol:chloroform:isoamyl (25:24:1) to each tube.
31. Mix vigorously and spin at >16,000 × *g* for 5 min at room temperature.
32. Transfer the aqueous phase (around 200 μl) to a fresh 1.5 ml Eppendorf tube, taking care not to touch the interphase.
33. For each tube, add 200 μl of chloroform:isoamyl (24:1).
34. Mix vigorously and spin at >16,000 × *g* for 5 min at room temperature.
35. Transfer the aqueous phase (around 200 μl) to a fresh 1.5 ml Eppendorf tube, taking care not to touch the interphase.
36. For each tube, add 20 μl of 3 M sodium acetate pH 5.5, 4 μl of GlycoBlue, and subsequently 1 ml of 100 % ethanol.
37. Mix vigorously and precipitate overnight at −80 °C.
38. The next day, spin tubes at 4 °C at >16,000 × *g* for 30 min. The blue pellet contains the RNA. Remove ethanol without touching the pellet.
39. For each sample, add 1 ml of 75 % ethanol and mix by inversion.
40. Spin tubes at 4 °C at >16,000 × *g* for 15 min. Remove ethanol without touching the pellet and repeat 75 % ethanol wash step once (for a total of two times).
41. Remove ethanol and dry blue RNA pellets in open fume hood, taking care not to over-dry them as this interferes with resuspension.
42. Resuspend RNA pellets in 20 μl of nuclease-free water and place on ice. The MRE RNA is now ready for RNA-seq.

3.10 IMPACT-seq Library Prep and Analysis

For IMPACT-seq library preparations, the use of small RNA kits is strongly recommended, and rRNA removal is discouraged. For IMPACT-seq on the Illumina platform, we prepared the library with the NEBNext small RNA library preparation kit with Illumina-specific primers, following the manufacturer's protocol. At the

size-selection step of the protocol that is generally performed with gel extraction, we select for an insert size of between 20 and 60 nt.

The bioinformatics analysis workflow for IMPACT-seq on the Illumina platform as previously published is as follows [13]:

1. Screen pre-alignment and post-alignment libraries for quality, specificity of mapping and containment sequences using FASTQC, RSeQC, and RNA-SeQC.
2. Trim low-quality bases, homopolymer sequences, and sequences matching the first 13 bases, and the reverse complement of the adapter sequences for Illumina using cutadapt, and discard trimmed reads smaller than 20 nt.
3. Align trimmed reads to the latest build of the human genome with Tophat, keeping only uniquely mapped reads with ≦2 mismatches.
4. These steps can be performed using the bipy (http://github.com/roryk/bipy) and bcbio-nextgen (http://github.com/chapmanb/bcbio-nextgen) automated sequencing analysis pipelines.
5. Call peaks for both control and experimental miRNA IMPACT-seq samples using CLIPper with the following parameters: --poisson-cutoff=0.05 --superlocal --max_gap=0 --processors=8 -b $file -s hg19 -o $file.
6. Identify MREs by requiring that each peak have ≥5 reads in the experimental miRNA sample with at least twice as many normalized reads in the experimental miRNA sample as control.

3.11 Analysis and Validation of MREs

The analysis of putative IMPACT-seq MREs can be accomplished by cloning a random set of these sequences into psiCHECK2, and performing 3′UTR luciferase assays [12–14].

1. For each IMPACT-seq MRE sequence, synthesize a phosphorylated, double-stranded DNA fragment, with a 5′ XhoI and a 3′ NotI restriction enzyme overhang sequence.
2. Digest psiCHECK2 with XhoI and NotI, and dephosphorylate cut vector with CIP.
3. Ligate dsDNA in **step 1** with cut vector in **step 2**, and transform into competent cells.
4. Pick colonies and perform PCR screening for positive clones with the Phusion kit, using a universal forward primer targeting the Renilla luciferase (ACCCTGGGTTCTTTTCCAAC), and a specific reverse primer that corresponds to the 3′–5′ sequence of the MRE.
5. The positive clones can then be sequenced using the universal forward primer to confirm the exact sequence of MRE that is cloned.

6. A positive control MRE (the reverse complement of the experimental miRNA sequence) and a few negative control MREs (a random sequence or known sequences that are not targets of the experimental miRNA) can also be cloned into psiCHECK2 in this way.
7. These psiCHECK2-MRE vectors are then transfected into the cell line of interest in 24-well plates, with either the control miRNA or the experimental miRNA using the protocol optimized above in Subheading 3.1.
8. At 48 h after transfection, luciferase activity is assayed with the Dual Luciferase Reporter kit and a plate reader.
9. The Renilla luciferase value (where the MRE is cloned) is normalized by the Firefly luciferase value, and this is calculated for both the control and experimental miRNA transfections. The ratio of experimental versus control value for each MRE is then compared to both negative and positive MRE control ratios to determine if the putative IMPACT-seq MRE is functional.

If more stringent validation is required, MRE sequences with a few point mutations of the expected miRNA-binding residues can be cloned into psiCHECK2 and assayed as above, with the expectation that these mutations will disrupt miRNA binding.

3.12 Bioinformatics Analysis of Target Genes

1. miRNAs regulate biological processes by targeting multiple genes involved in similar pathways [4]. To decipher biological functions of a miRNA, targets identified by the pulldown method can be analyzed by gene ontology, interactome and/or pathway analysis software, for example Ingenuity Pathway Analysis (IPA, www.ingenuity.com) and Database for Annotation, Visualization and Integrated Discovery (DAVID, david.abcc.ncifcrf.gov) [12–14].
2. Gene ontology, interactome, and pathway analysis can also be conducted on genes that are downregulated after overexpression of the miRNA. This analysis takes into account genes regulated both directly and indirectly by the miRNA.
3. For pathway analysis of both pulldown targets and downregulated genes, we recommend the use of IPA (www.ingenuity.com). For the two gene lists, first connect all genes that are directly related with the default settings, as defined by the curated IPA database, without selecting the optional predicted miRNA interactions. Second, conduct a Core analysis with the default IPA settings, to generate IPA scores for the top associated network functions and significantly enriched molecular functions for both lists. Finally, identify network and molecular functions that are common in both pulldown targets and downregulated genes. This analysis can be done using different stringency cutoffs to select the target genes. The most important functions will persist with more stringent cutoffs since the

most highly regulated genes are generally most enriched in the pulldown.

4. A useful addition to gene ontology and pathway analyses is to search for over-represented transcription factors predicted to bind to the promoter regions upstream of target genes. This analysis, which can be done using TRANSFAC, a well-curated knowledge base of eukaryotic transcription factors (www.gene-regulation.com), can shed light on miRNA function based on the known functions of the transcription factors that transcriptionally regulate the miRNA target genes [13]. Combining all these methods should result in a set of hypotheses about the set of biological functions of the mRNA that can direct experimental efforts into fruitful investigations that avoid unproductive searching. This streamlined approach should be useful for uncovering the hidden meaning of large genome-wide datasets more generally.

4 Notes

1. We recommend the use of miRNA mimics from Dharmacon, which are biotinylated at the 3′ end of the active strand, and are designed with proprietary modifications to both the active and passive strands to improve loading into the miRISC. Trials with siRNA-like biotinylated dsRNA with and without bulges did not work as well.
2. Optimizing miRNA transfection efficiency for the cell line of interest is crucial for the success of this protocol. This may require testing other transfection lipids or using nucleoporation to transfect cells that are ordinarily more difficult to transfect (such as nonadherent cells, especially hematopoietic cells and lymphocytes). We strongly recommend optimizing transfection to maximize both the number of transfected miRNAs per cell and the uniformity of transfection. One way to test the efficiency of any transfection protocol is to monitor mRNA levels of known miRNA targets 24 or 48 h after transfection via real-time PCR. For miRNAs without any previously validated targets, one approach is to test the top conserved or highest scoring genes predicted by algorithms (for example TargetScan) as positive controls. Another approach is to test the effect of the transfected miRNA on luciferase activity in cells transiently transfected with a 3′UTR dual-luciferase reporter, such as psiCHECK2 [12–14], containing an antisense sequence to the miRNA active strand. Transfection conditions that cause >90 % reduction in expression of known targets or luciferase activity are ideal. Alternatively, one may optimize transfection of fluorescently labeled miRNAs (or siRNAs) using flow cytometry analysis of the miRNA tag to measure miRNA internalization. Low transfection may impair the ability to detect pulled down targets (*see* **Note 11**). In addi-

tion, we recommend performing pulldown experiments in cell lines with a moderate level of expression of your miRNA of interest. This ensures that the transfected biotinylated miRNA mimics will not be diluted out by high levels of the endogeneous miRNA, enabling efficient capture of targets. Performing pulldowns in such cell lines will also allow for identification of biologically relevant targets, as both the miRNA of interest and its targets are expressed in the same system.

3. The length of time between transfection and cell lysis also needs to be optimized for each cell line and miRNA. The optimal timing balances the opposing considerations of increasing binding of the miRNA to miRISC and target RNAs with the miRNA-mediated accelerated degradation of the target mRNA. In our experience, overnight or 20 h is usually ideal. However, for some cells or miRNAs the optimal time, determined using the systems described in **Note 2**, needs to be shortened to optimize the pulldown. To optimize the timing, the fold enrichment of known target mRNAs should be assessed by real-time PCR after pulldowns at 4, 8, 12, 16, and 20 h.
4. It is important to resuspend cells to a single-cell state as much as possible before the addition of Lysis Buffer Plus for efficient lysis.
5. As TRIzol LS is used to precipitate RNA, be cautious about the carry-over of guanidine salts to the RNA prep. To minimize the effect of these salts (which can inhibit PCR reactions), we use only a small amount of RNA (2 μl of the total 50 μl of RNA) for each 30 μl cDNA reaction.
6. Performing the pulldowns with multiple miRNAs in parallel can be useful as these additional miRNAs can be used to optimize the transfection conditions and they can also serve as controls for specificity. Target genes that are enriched for binding to multiple miRNAs may bind nonspecifically.
7. The cutoffs used to define the list of miRNA-regulated genes are in some ways arbitrary. It is unknown in any context how many of the genes that a miRNA can potentially regulate are physiologically important for miRNA function. It is important to keep in mind that the pulldown identifies potentially regulated genes that are isolated under conditions of miRNA overexpression. We have found that the fold enrichment in the pulldown correlates with the extent of gene downregulation by the miRNA [12–14]. Since the most highly enriched mRNAs may be those that are most strongly regulated by the miRNA, using more stringent cutoffs can reduce the gene lists to a more manageable size and may focus further experimentation on genes most likely to be important, however, at the risk of losing important targets. The overlap of the set of genes down-regulated after miRNA overexpression from microarray or RNA-seq data with the set of genes enriched in the pull-

down can also be used to define a more focused candidate target gene set, testing a variety of cutoffs for each component to produce different sized gene lists. One way to determine the best cutoffs is to compare plots of cumulative changes induced by the miRNA in mRNA levels of the gene sets defined by different cutoffs (with statistical analysis using the Kolmogorov-Smirnov test) [13]. Increases in the stringency of the cutoff that reduce the gene list size, but do not significantly change overall gene downregulation, reduce the size of the gene list, but are unlikely to select better targets.

8. In choosing control genes in the analysis of pulldown or knockdown results with real-time PCR, we recommend that multiple negative control housekeeping genes (such as *GAPDH*, *ACTB* and *SDHA*) be tested. Some miRNAs target one or more of the commonly used housekeeping genes. Genes, whose expression does not vary after transfection of the miRNA, should be chosen for normalizing the PCR results.
9. It is useful to include positive controls for known miRNA targets in real-time PCR analyses, if possible. This is a good indication of the efficiency of each pulldown experiment.
10. Since most miRNA targets identified by enrichment in the pulldown are also downregulated by over-expression of the miRNA [12–14], we recommend validating novel miRNA targets by examining changes in their expression after miRNA over-expression.

Although multiple studies from various groups have identified miRNA targets with this technique [12–19], a recent report suggests that biotinylated miRNAs do not associate with AGO, based on analysis of miR-27 [21]. This finding seems to contradict our finding that AGO depletion removes 70–80 % of target gene mRNAs from the miR-522 pulldown [13]. One possible explanation for their negative result could be their low level of transfection of the miRNA mimic. This reinforces our emphasis on the importance of high transfection efficiency (*see* **Note 2**). However, another potential explanation is that the biotin tag may interfere with miRISC/AGO binding of some biotinylated miRNAs. In particular, we were unable to get the pulldown to work well for miR-21 in some cells. This could be a sequence-specific or even a context-specific problem, since posttranslational modifications of AGO can affect miRNA binding [22].

Acknowledgements

We thank Jingmin Jin and Larry McReynolds for their sequencing expertise, Rory Kirchner, Oliver Hofmann and Winston Hide for their bioinformatics expertise, and members of the Lieberman lab for critical discussions. S.M.T. was supported by the Department of Defense (DOD) Breast Cancer Research Program (BCRP).

References

1. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233
2. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20
3. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell 126:1203–1217
4. Thomas M, Lieberman J, Lal A (2010) Desperately seeking microRNA targets. Nat Struct Mol Biol 17:1169–1174
5. Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460:479–486
6. Loeb GB, Khan AA, Canner D, Hiatt JB, Shendure J, Darnell RB, Leslie CS, Rudensky AY (2012) Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. Mol Cell 48:760–770
7. Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW (2010) Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. Nat Struct Mol Biol 17:173–179
8. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141
9. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665
10. Grosswendt S, Filipchyk A, Manzano M, Klironomos F, Schilling M, Herzog M, Gottwein E, Rajewsky N (2014) Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell 54:1042–1054
11. Keene JD, Komisarow JM, Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. Nat Protoc 1:302–307
12. Lal A, Thomas MP, Altschuler G, Navarro F, O'Day E, Li XL, Concepcion C, Han YC, Thiery J, Rajani DK et al (2011) Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. PLoS Genet 7:e1002363
13. Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W, Lieberman J (2014) Sequencing and systems analysis of captive target transcripts identifies primate-specific miR-522 as an inducer of mesenchymal transition. Cell Rep 8:1225–1239
14. Perdigao-Henriques R, Petrocca F, Altschuler G, Thomas MP, Le MT, Tan SM et al. miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. Oncogene 2015. Advanced doi: 10.1038/onc.2015.69.online publicaton
15. Orom UA, Nielsen FC, Lund AH (2008) MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell 30:460–471
16. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E et al (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol 12:R126
17. Kang H, Davis-Dusenbery BN, Nguyen PH, Lal A, Lieberman J, Van Aelst L, Lagna G, Hata A (2012) Bone morphogenetic protein 4 promotes vascular smooth muscle contractility by activating microRNA-21 (miR-21), which down-regulates expression of family of dedicator of cytokinesis (DOCK) proteins. J Biol Chem 287:3976–3986
18. Le MT, Shyh-Chang N, Khaw SL, Chin L, Teh C, Tay J, O'Day E, Korzh V, Yang H, Lal A et al (2011) Conserved regulation of p53 network dosage by microRNA-125b occurs through evolving miRNA-target gene pairs. PLoS Genet 7:e1002242
19. Martin HC, Wani S, Steptoe AL, Krishnan K, Nones K, Nourbakhsh E, Vlassov A, Grimmond SM, Cloonan N (2014) Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs. Genome Biol 15:R51
20. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505:344–352
21. Guo YE, Steitz JA (2014) 3′-Biotin-tagged microRNA-27 does not associate with Argonaute proteins in cells. RNA 20:985–988
22. Meister G (2013) Argonaute proteins: functional insights and emerging roles. Nat Rev Genet 14:447–459

Chapter 14

Identification of miRNA-Target RNA Interactions Using CLASH

Aleksandra Helwak and David Tollervey

Abstract

We present a detailed protocol for the experimental identification of miRNA-target RNA interaction sites using *c*ross-linking, *l*igation, *a*nd *s*equencing of *h*ybrids (CLASH). The basis of the technique is the purification of UV-stabilized Argonaute (AGO)–RNA complexes assembled in living cells, with subsequent ligation of AGO-associated RNA-RNA duplexes to form chimeric RNAs. Following cDNA synthesis, DNA library preparation and high-throughput sequencing, interacting RNA molecules are unambiguously identified as chimeric reads in bioinformatic analysis of sequencing data. CLASH potentially recovers any RNA duplex that is bound by RNA-binding protein, so modified approaches would be suitable for the identification of many other inter- and intramolecular RNA-RNA interactions. Since CLASH analysis is independent of bioinformatic predictions it allows the identification and analysis of RNA targeting rules in an unbiased way.

Key words CLASH, CLIP, RNA-RNA interactions, Protein-RNA interactions, Argonaute, microRNA, miRNA target identification, RNA cross-linking, UV cross-linking

1 Introduction

Advances in next-generation sequencing technology are revealing the transcriptome in ever-increasing detail, bringing new understanding of the complexity and extent of RNA-based regulation [1, 2]. RNAs frequently act within large, multi-subunit complexes (RNPs) and perform multiple functions, carrying the genetic information (viral genomes, mRNA), performing enzymatic roles (ribosomes, spliceosomal snRNAs, ribozymes), serving as a structural scaffold for protein complexes (lncRNAs), or guiding proteins to their site of action (miRNAs, snoRNAs). RNA molecules are capable of performing such variable roles due to their structural flexibility, both in the mechanical and evolutionary sense, and ability to form numerous interactions with other RNA molecules and proteins.

miRNAs are small noncoding RNA molecules involved in the posttranscriptional regulation of gene expression. They form

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_14,

complexes with AGO proteins, called RNA-induced silencing complexes (RISCs), and guide them to their target sites. Currently most miRNA target sites on mRNAs are predicted using bioinformatics, and validated by experimental miRNA depletion data. Continued development of bioinformatic techniques has yielded increasingly reliable data, but these are still largely confined to the identification of seed-based interactions in 3′ UTRs [3]. These are, on average, the most significant functional interactions, at least at the level of mRNAstability [4]. However, significant non-seed interactions have been functionally characterized [5, 6], and it is clear that, due to protein binding and other structural constraints [7], not all potential base-pairing sites will actually be accessible in any specific cell type. This is likely to particularly be the case in systems with complex post-transcriptional regulation such as neurons. Moreover, the role of noncoding RNAs (ncRNAs) as regulators of gene expression and miRNA function is increasingly evident. Bioinformatic prediction of ncRNA-miRNA interactions that potentially regulate the function of either the ncRNA or miRNA are highly problematic at present. These considerations make an unbiased, experimental approach to miRNA target identification of value.

This chapter describes the experimental technique of *cross-linking*, *l*igation *a*nd *s*equencing of *h*ybrids (CLASH) that allows the high-throughput identification of both protein-RNA and RNA-RNA interactions, which are formed in living cells in proximity to the binding site for a selected protein. Methodologically CLASH combines two experimental strategies: (*see* Fig. 1) (1) In vivo UV crosslinking and isolation of protein-RNA complexes, similar to CLIP [8]/CRAC [9] techniques. (2) Ligation of RNA molecules that interact within RNPs to form chimeric RNAs. An important part of CLASH is the bioinformatic analysis that identifies interacting molecules from the high-throughput sequence data. The bioinformatic pipeline is briefly outlined in Fig. 3 and has previously been described in detail in [10, 11].

The detailed experimental protocol reported here is specifically adjusted for recovery of miRNA interactions, and has been applied to the characterization of miRNA interactome in human HEK293 cells [10]. In outline; the first step in CLASH is expression of the tagged AGO protein. Living cells are briefly UV cross-linked to stabilize endogenous AGO-RNA complexes and miRNA–target RNA interactions within these complexes are maintained by base-pairing. RNPs are extensively purified under protein-denaturing conditions, and associated RNAs are trimmed with RNases to about 25–30 nucleotide long AGO protected fragments. Interacting RNA molecules are then ligated together to form chimeric RNAs. In subsequent steps RNA is isolated and converted into the DNA library ready for high throughput sequencing. The CLASH protocol has been published in [12] and a related approach has recently been described by Rajewsky's group [13].

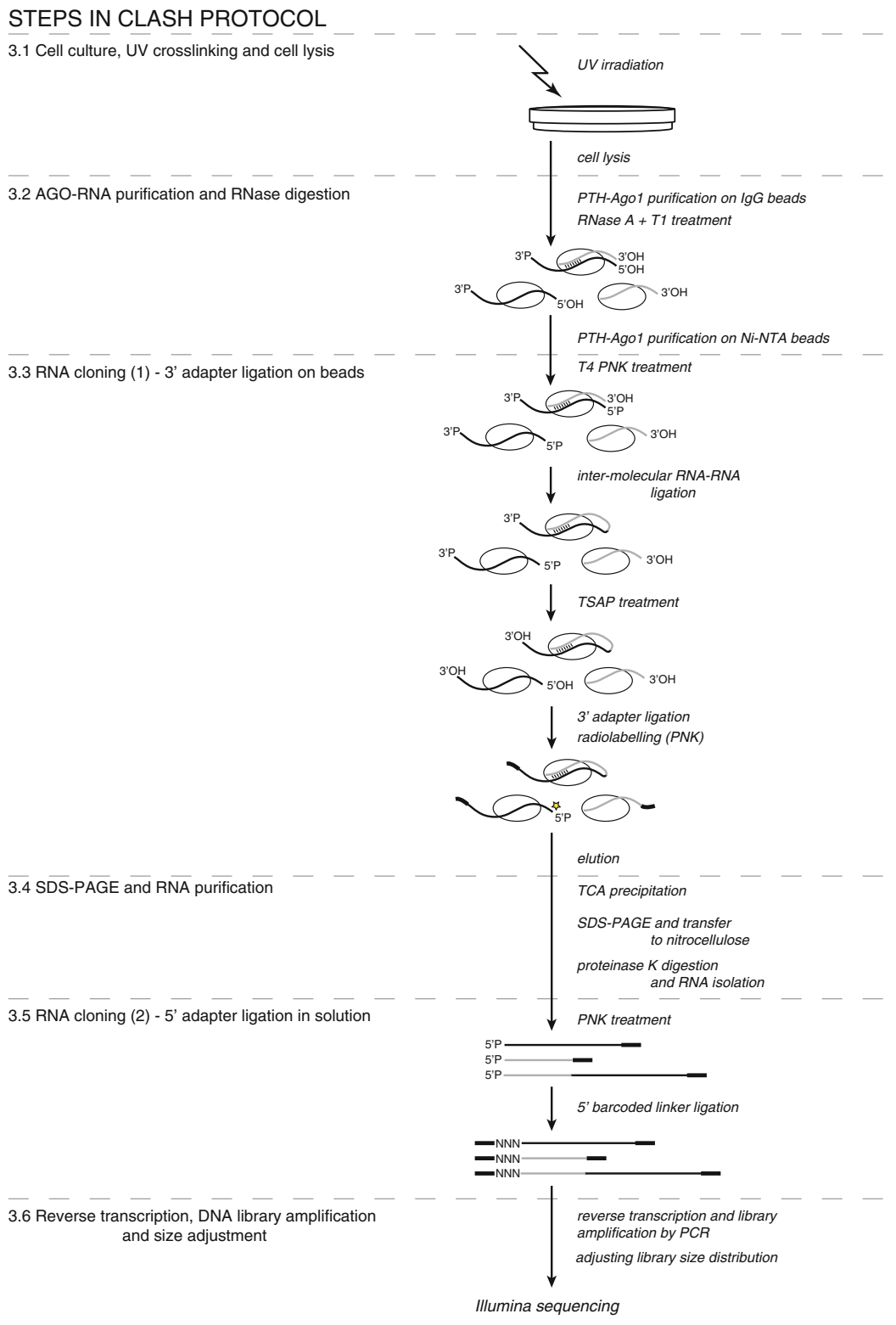

Fig. 1 Outline of the AGO-CLASH experimental protocol. Adapted from [10]

2 Materials

All the reagents should be free of RNases and prepared with deionized water. DEPC treatment is not essential. To keep the buffers as clean as possible, adjust the pH by taking small aliquots for measurements. Filter-sterilize all stock solutions following preparation and store at 4 °C. Where required, add β-mercaptoethanol and protease inhibitors to the buffers shortly before use.

2.1 Cell Culture Components

1. Cells expressing AGO protein fused to an N-terminal PTH tag (Protein A—TEV site-6×His) and the parental cell line not expressing the tagged protein for use as a negative control (*see* **Note 1**).
2. 150 mm culture dishes.
3. Growth medium adequate for the cells.
4. Dulbecco's PBS.

2.2 Buffer Compositions and Other Solutions

1. Lysis buffer: 50 mM Tris–HCl (pH 7.8), 300 mM NaCl, 1 % NP-40 (vol/vol), 5 mM EDTA (pH 8.0) and 10 % glycerol (vol/vol), 5 mM β-mercaptoethanol, complete Protease inhibitors, EDTA-free (Roche Applied Science).
2. LS (low salt) IgG buffer: 50 mM Tris–HCl (pH 7.8), 300 mM NaCl, 0.5 % NP-40 (vol/vol), 2.5 % glycerol, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol.
3. HS (high-salt)-IgG buffer: 50 mM Tris–HCl (pH 7.8), 800 mM NaCl, 0.5 % NP-40 (vol/vol), 2.5 % glycerol, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol.
4. PNK-WB (wash buffer): 50 mM Tris–HCl (pH 7.8), 50 mM NaCl, 0.5 % NP-40 (vol/vol), 10 mM $MgCl_2$, 5 mM β-mercaptoethanol.
5. PNK buffer (5× reaction buffer): 250 mM Tris–HCl (pH 7.5), 250 mM NaCl, 2.5 % NP-40 (vol/vol), 50 mM $MgCl_2$, 50 mM β-mercaptoethanol. Aliquot and store at –20 °C.
6. Ni-WBI (wash buffer I): 50 mM Tris–HCl (pH 7.8), 300 mM NaCl, 0.1 % NP-40 (vol/vol), 10 mM imidazole (pH 8.0), 6 M guanidine hydrochloride, 5 mM β-mercaptoethanol. Prepare fresh before each experiment and protect from light.
7. Ni-WBII (wash buffer II): 50 mM Tris–HCl (pH 7.8), 300 mM NaCl, 0.1 % NP-40 (vol/vol), 10 mM imidazole (pH 8.0), 5 mM β-mercaptoethanol. Protect the buffer from light.
8. Ni–EB (elution buffer): 50 mM Tris–HCl (pH 7.8), 50 mM NaCl, 0.1 % NP-40 (vol/vol), 150 mM imidazole (pH 8.0), 5 mM β-mercaptoethanol. Protect the buffer from light.

9. Proteinase K buffer: 50 mM Tris–HCl (pH 7.8), 50 mM NaCl, 0.1 % NP-40 (vol/vol), 10 mM imidazole (pH 8.0), 1 % SDS (wt/vol), 5 mM EDTA (pH 8.0), 5 mM β-mercaptoethanol. Aliquot and store at –20 °C.
10. PCI: phenol-chloroform-isoamyl alcohols are mixed in volume proportions 25:24:1. Store at 4 °C. Discard if the color of the mixture changes.
11. 10× TBE buffer: 890 mM Tris base, 890 mM boric acid, 20 mM EDTA. Store at room temperature.
12. Trichloroacetic acid (TCA), 100 % (wt/vol).
13. Guanidine hydrochloride (GuHCl) powder.
14. Acetone.
15. Methanol.
16. Ethanol.
17. 3 M NaOAc (sodium acetate) pH 5.5.

2.3 Enzymes and Enzymatic Reaction Components (see **Note 2***)*

1. RNace-IT (Agilent) RNase A+T1, working stock prepared by diluting 1:20 in water, store for a long term at 4 °C (*see* **Note 3**).
2. ATP, 100 and 10 mM solutions in water, aliquot and store at –20 °C, avoid repeated freezing and thawing.
3. RNase inhibitor, e.g., RNasin (Promega).
4. T4 PNK, T4 Polynucleotide Kinase (New England BioLabs).
5. ^{32}P-γ-ATP (6000 Ci/mmol).
6. T4 RNA ligase 1 (New England BioLabs).
7. TSAP, Thermosensitive Alkaline Phosphatase (Promega).
8. PEG 8000 supplied with T4 RNA ligase 2 truncated, K227Q.
9. T4 RNA ligase 2 truncated, K227Q (New England BioLabs).
10. Proteinase K (Roche Applied Science), prepare 20 mg/ml stock in deionized water, aliquot and store at –20 °C.
11. T4 RNA ligase reaction buffer, 10× (supplied with T4 RNA ligase 1).
12. SuperScript III Reverse Transcriptase (Life Technologies).
13. dNTPs, 2.5 mM.
14. First strand buffer, 5× (supplied with SuperScript III Reverse Transcriptase).
15. 0.1 M DTT (supplied with SuperScript III Reverse Transcriptase).
16. RNase H (New England BioLabs).
17. TaKaRa LA Taq (Clontech).
18. LA PCR Buffer ll (Mg^{2+} plus), 10× supplied with TaKaRa LA Taq.

Table 1
List of Illumina-compatible 5′ and 3′ adapters and RT and PCR primers

ID	Oligonucleotide sequence		
Illumina-compatible barcoded 5′ adapter, *C* = 100 μM			
	Common sequence:	Random barcode	Sample barcode:
L5Aa	5′-invddT-ACACrGrArCrGrCrUr	rNrNrN	rUrArArGrC-3′OH
L5Ab	CrUrUrCrCrGrArUrCrU		rArUrUrArGrC-3′OH
L5Ac			rGrCrGrCrArGrC-3′OH
L5Cc			rArCrTrCrArGrC-3′OH
L5Cd			rGrArCrTrTrArGrC-3′OH
Illumina-compatible 3′ adapter, *C* = 10 μM			
miRCat-33	AppTGGAATTCTCGGGTGCCAAG/ddC/		
Primers for reverse transcription and PCR amplification of the DNA library, *C* = 10 μM			
miRCat-33 RT primer	CCTTGGCACCCGAGAATT		
PE_miRCat_PCR	CAAGCAGAAGACGGCATACGA GATCGGTCTCGGCATTCCTGGCCTTGGCACCCGAGAATTCC		
P5	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCT		

19. Illumina compatible adapters and RT and PCR primers are included in Table 1. miRCat-33 Conversion Oligos Pack (miRCat-33 adapter and miRCat-33 RT primer; Integrated DNA Technologies; IDT), other oligonucleotides synthesized by custom order (IDT). After dissolving, aliquot adapters and store at −80 °C (*see* **Note 4**).

2.4 Laboratory Equipment

1. Humidified 37 °C, 5 % CO_2 incubator.
2. Cell culture hood.
3. Cross-linker—Stratalinker 1800 (Stratagene) with UV bulbs, $\lambda = 254$ nm or equivalent.
4. Magnetic rack for 15 ml conical tubes.
5. Magnetic rack for microcentrifuge tubes.
6. Rotating wheel for microcentrifuge tubes.
7. Thermoblock with shaking.
8. Refrigerated centrifuge for conical 15 ml tubes.
9. Refrigerated benchtop centrifuge.
10. SDS-PAGE tank XCell SureLock Mini-Cell for NuPAGE gels.
11. Wet-transfer apparatus, Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) or equivalent (*see* **Note 5**).
12. Thermocycler for PCR amplification of the DNA library.

13. Agarose gel electrophoresis cell.
14. Gel scanner attached to the printer, able to print scan it its original size.
15. Qubit Fluorimeter (Life Technologies).
16. Film developer.
17. Vortex.

2.5 Other Labware and Consumables

1. Small tray that can be filled with ice and fitted into the cross-linking chamber.
2. Cell scrapers.
3. Filter units for buffers sterilization with pore size 0.2 μm.
4. Dynabeads M-270 epoxy (Life Technologies) coated in advance with total IgG from rabbit serum according to the protocol [14] (*see* **Note 6**).
5. Ni-NTA Superflow Agarose beads (Qiagen) or similar 6×His-binding resin.
6. Spin Columns with snap-caps (Thermo Scientific).
7. Bovine serum albumin (BSA) dissolved in water at 10 mg/ml.
8. Pre-stained protein standard: SeeBlue Plus2 (Life Technologies) or equivalent.
9. NuPAGE LDS Sample Buffer, 4× (Life Technologies).
10. NuPAGE 4–12 % polyacrylamide Bis-Tris gels (Life Technologies).
11. NuPAGE MOPS SDS running buffer (Life Technologies).
12. Nitrocellulose membrane Hybond-C Extra membrane (GE Healthcare) or equivalent.
13. NuPAGE Transfer Buffer (Life Technologies).
14. Phosphorescent rulers for autoradiography.
15. GlycoBlue (Life Technologies) or glycogen for RNA precipitation.
16. Kodak BioMax MS Autoradiography Film or equivalent.
17. PCR purification kit with low elution volumes, e.g., MinElute PCR purification kit (Qiagen).
18. MetaPhor high resolution agarose (Lonza).
19. SYBRSafe (Life Technologies).
20. 50 bp DNA ladder, e.g., GeneRuler 50 bp DNA ladder (Thermo Scientific).
21. DNA Loading dye, e.g., 6× DNA Loading dye (Thermo Scientific) or equivalent.
22. DNA Gel extraction kit with low elution volumes e.g. MinElute Gel extraction kit (Qiagen).

23. Transparency film.
24. Scalpels.
25. Qubit dsDNA HS Assay Kit (Life Technologies).

3 Methods

Unless specified otherwise, keep the samples on ice and use ice-cold wash buffers.

3.1 Cell Culture, UV Cross-Linking, and Cell Lysis

1. *Cell culture*. Seed the cells expressing tagged AGO protein and control cells on 150 mm diameter culture dishes and grow them in adequate growth medium until they approach confluence. Use cells from four dishes per CLASH analysis (*see* **Note 7**).
2. *UV cross-linking*. Tip the dish and discard cell medium. Rinse the cells briefly with 10 ml room-temperature DPBS, remove the buffer and place the dish without the lid on the ice tray. Transfer the tray into the Stratalinker and irradiate the cells at $\lambda = 254$ nm with total energy output set to 400 mJ/cm^2. Directly after irradiation add 2.5 ml lysis buffer to each dish (*see* **Note 8**).
3. *Cell lysis*. Incubate the culture dish on ice for 5 min or until all the dishes are ready for further steps. Scrape cell remains from the dishes and transfer to the conical tubes. Centrifuge the lysates at 6400 × *g* for 30 min at 4 °C and collect supernatants (*see* **Note 9**).
4. Prepare western blot "IgG-input" sample by boiling 15 µl lysate with 5 µl NuPAGE LDS sample buffer for 5 min (*see* **Note 10**). Store the samples at −20 °C for further use at Subheading 3.2, **step 7**. Use the entire sample per gel lane.

3.2 AGO-RNA Purification and RNase Digestion

1. *AGO-RNA purification on IgG*. Wash IgG-coated Dynabeads three times with 5 ml ice-cold PBS (*see* **Note 11**). Combine the beads with fresh or freshly thawed cell lysates (approximate volume 10 ml), using 20 mg of beads per sample. Incubate with rotation for 40 min at 4 °C.
2. Place the tubes in the magnetic rack, wait until all the beads settle at the side of the tube, and gently remove the supernatant, saving 15 µl of supernatant for western blot ("IgG flow through" samples: boil 15 µl supernatant with 5 µl NuPAGE LDS sample buffer for 5 min; store the samples at −20 °C for further use at **step 7**. Use the entire sample per gel lane). Wash the beads: 2 × 10 ml briefly with ice-cold LS-IgG buffer, 2× with 10 ml ice-cold HS-IgG buffer incubating samples with rotation for 5 min at 4 °C during each wash and finally once briefly with 10 ml ice-cold PNK-WB buffer. Gently resuspend the beads in 1 ml PNK-WB buffer and transfer them to 1.5 ml

microcentrifuge tube. Pellet the beads in the magnetic rack and remove supernatant.

3. *RNase digestion*. Prepare RNase digestion buffer (*see* **Note 2**): for each sample mix 1 μl (0.5 U) of RNace-IT dilution and 0.5 ml room temperature PNK-WB (*see* **Note 3**). Incubate the samples with RNase digestion buffer for 7 min at 20 °C, tapping the samples gently every 30 s to prevent beads settling. Place all the samples on ice, pellet the beads quickly, and remove the supernatant.
4. Elute AGO-RNA complexes from IgG-Dynabeads with 0.5 ml Ni-WBI. Incubate the beads suspension for 10 min at room temperature with rotation, then pellet the beads and collect the supernatant. Elute twice more, once with 0.5 ml and once with 1 ml Ni-WBI. Pool all the eluates together.
5. Prepare western blot sample "Ni-NTA input" with the following procedure: 10 μl eluate from previous step + 90 μl water + 100 μl 10 % TCA, incubate for 30 min on ice, follow the procedure for TCA precipitation described in Subheading 3.4, **steps 1–3**. Resuspend the pellets in 15 μl water, add 5 μl NuPAGE LDS sample buffer and boil for 5 min. Use the entire sample per gel lane (*see* **Note 13**). To control for the elution efficiency prepare western blot sample "IgG beads." Wash the IgG Dynabeads with 1 ml PNK-WB, resuspend them in 1 ml PNK-WB, take 15 μl beads suspension, add 135 μl water and 50 μl NuPAGE LDS sample buffer, boil for 5 min. Use 20 μl per gel lane.
6. *AGO-RNA purification on Ni-NTA*. Prepare 40 μl Ni-NTA beads per sample (approximate settled beads volume) (*see* **Note 14**). Equilibrate all the beads together by washing twice with 1 ml with Ni-WBI. Combine the beads with eluates from **step 4** and incubate with rotation for 2 h at 4 °C.
7. Pellet the beads by centrifugation at 1000 × *g* for 10 s at 4 °C and gently remove supernatant. Save 10 μl supernatant for western blot sample "Ni-NTA flow through," follow the same protocol as for Ni-NTA input (**step 5**). At this stage western blot with all the samples collected during AGO-RNA purification can be performed (*see* **Note 15**).
8. Wash the beads twice with 1 ml ice-cold Ni-WBI, incubating samples with rotation for 10 min at 4 °C during each wash (*see* **Note 16**).
9. Transfer the beads to the small snap cap spin columns and place the columns in the 1.5 ml microcentrifuge tubes or lidless collection tubes. Keep the samples cold (*see* **Note 17**).
10. Wash the beads on the column twice with 0.75 ml Ni-WBII and 3× with PNK-WB (*see* **Note 18**). Dry the beads by closing the lid of the column.

3.3 RNA Cloning (1): 3′ Adapter Ligation on Beads

1. *T4 PNK phosphorylation of the RNA 5′ ends.* Prepare 80 μl T4 PNK phosphorylation mixture per sample: 57 μl water, 16 μl 5× PNK buffer, 0.8 μl 100 mM ATP (final concentration 1 mM), 2 μl RNAsin (final concentration 1 U/μl), and 4 μl T4 PNK (final concentration 0.5 U/μl). Close the column bottom with the supplied white snap cap and add phosphorylation mixture to each sample (*see* **Note 19**). Incubate samples for 2.5 h at 20 °C.
2. Wash the beads as follows: 2× with 0.5 ml Ni-WBI, 2 × 0.75 ml ice-cold Ni-WBII and 3 × 0.75 ml PNK-WB, discard the washes (*see* **Note 20**).
3. *Intermolecular RNA-RNA ligation*. Prepare 160 μl RNA ligation mixture per sample: 118.4 μl water, 32 μl 5× PNK buffer, 1.6 μl 100 mM ATP (final concentration 1 mM), 4 μl RNAsin (final concentration 1 U/μl), and 4 μl T4 RNA ligase 1 (final concentration 0.25 U/μl). Close the column bottom with the supplied snap cap, add RNA ligation mixture, and incubate samples with gentle rotation overnight at 16 °C (*see* **Note 21**).
4. Wash the beads as in **step 2**.
5. *TSAP dephosphorylation of RNA 5′ and 3′ ends.* Prepare 80 μl dephosphorylation mixture per sample: 54 μl water, 16 μl 5× PNK buffer, 2 μl RNAsin (final concentration 1 U/μl), and 8 μl TSAP (final concentration 0.1 U/μl). Close the column bottom with the supplied snap cap, add dephosphorylation mixture and incubate samples for 45 min at 20 °C.
6. Wash the beads as in **step 2**.
7. *Ligation of 3′ sequencing adapter*. Prepare 80 μl ligation mixture per sample: 18 μl water, 16 μl 5× PNK buffer, 32 μl 25 % PEG 8000 (final concentration 10 %), 8 μl 10 μM miRCat-33 3′-adapter (final concentration 1 μM), 2 μl RNAsin (1 U/μl), 40 μl T4 RNA ligase 2, truncated, K227Q (final concentration 10 U/μl). Close the column bottom with the supplied snap cap, add ligation mixture, and incubate for 6 h at 16 °C.
8. Wash the beads as in **step 2**.
9. *Radioactive labelling of RNA*. Prepare 80 μl labelling mixture per sample: 55 μl water, 16 μl 5× PNK buffer, 3 μl ^{32}P-γ-ATP, 2 μl RNAsin (final concentration 1 U/μl), and 4 μl T4 PNK (final concentration 0.5 U/μl). Close the columns with the supplied snap caps, add labelling mixture, and incubate samples for 30 min at 37 °C (*see* **Note 22**).
10. Wash the beads multiple times with Ni-WBI until radioactivity of the flow-through buffer gets low (*see* **Note 23**).
11. Wash the beads as in **step 2**, and then dry the beads well by closing the lid two or three times.

12. *Elution of the AGO-RNA complexes from the beads.* Close the column with a snap cap, add 200 µl Ni-EB buffer and incubate the sample with mixing for 5 min at room temperature. Collect the supernatant. Repeat the elution twice more, once with 200 µl and once with 600 µl. Pool all the eluate fractions together (*see* **Note 24**).

3.4 SDS-PAGE and RNA Purification

1. *TCA precipitation.* To the eluate from previous step add 2 µg BSA and 200 µl 100 % TCA and incubate for 30 min on ice. Then pellet the AGO-RNA complexes by centrifugation with 20,000 × *g* (or max speed) for 30 min at 4 °C, discard supernatant (*see* **Note 25**).
2. Wash the pellets with 1 ml acetone pre-chilled in –20 °C, vortex the samples well and centrifuge for 10 min at 4 °C. Discard the supernatants. Repeat the wash procedure one more time (*see* **Note 26**).
3. Dry the pellets at room temperature (*see* **Note 27**).
4. Resuspend the pellets in 15 µl water and 5 µl 4× NuPAGE LDS sample buffer. Heat the samples for 10 min at 65 °C (*see* **Note 28**).
5. *SDS-PAGE.* Resolve the samples and a pre-stained protein ladder on the 4–12 % Bis-Tris NuPAGE gel with 1× NuPAGE MOPS SDS running buffer at constant voltage (100 V) (*see* **Note 29**). Proceed with the separation until the dye reaches the bottom of the gel.
6. *Transfer of AGO-RNA complexes.* Transfer protein-RNA complexes from the gel to the nitrocellulose membrane using cooled wet-transfer tank and 1× NuPAGE transfer buffer supplemented with 10 % methanol. Proceed with the transfer for 2 h at constant voltage (100 V) (*see* **Note 30**).
7. Retrieve the membrane and let it quickly air-dry. Wrap it in the cling film and attach the phosphorescent ruler. Expose the membrane onto the autoradiography film (*see* **Note 31**).
8. Develop the film and carefully align with the membrane using the signal from the phosphorescent ruler. Cut out the radioactive membrane fragments corresponding to the PTH-AGO1-RNA complexes from all the samples including the negative control using disposable scalpels. Follow the instruction in the Fig. 2a (*see* **Notes 32** and **33**).
9. *Proteinase K treatment.* Place cut out fragments of the nitrocellulose membrane in separate 1.5 ml microcentrifuge tubes (*see* **Note 34**). To each tube add 400 µl proteinase K buffer and 100 µg Proteinase K. Incubate the samples with gentle mixing for 2 h at 55 °C. Discard the membrane.

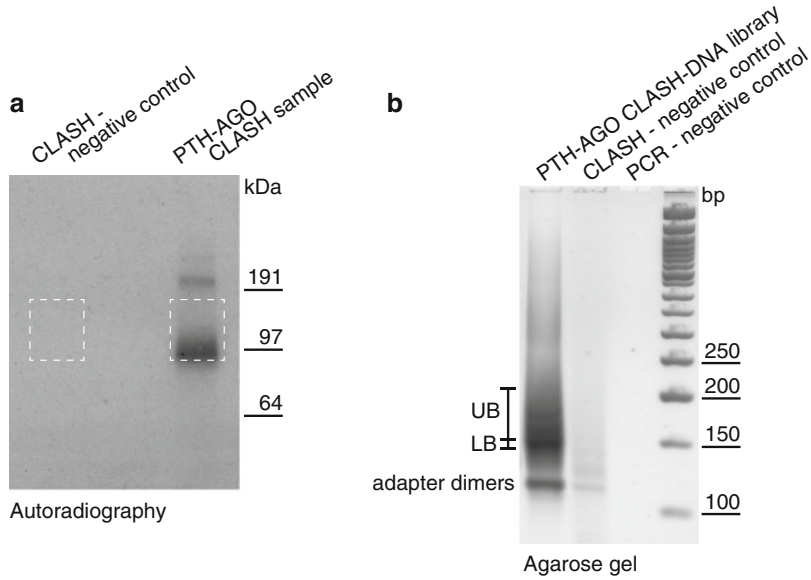

Fig. 2 Key steps in the AGO-CLASH protocol. (**a**) Autoradiogram indicating a difference in radioactivity arising from labeling of the RNA fragments, between an experimental sample and a negative control. *Dashed white frames* indicate fragments cut out of the nitrocellulose membrane at Subheading 3.4, **step 8**. (**b**) PCR amplification products resolved on the MetaPhor gel at Subheading 3.6, **step 9**. *Bars* at the side mark the gel fragments that were excised as upper band (UB) and lower band fractions (LB). DNA was isolated and both fractions were pooled together in desired ratio for the high-throughput sequencing. Adapted from [12]

10. *RNA extraction and precipitation*. To the supernatant add 50 μl 3 M NaOAc pH 5.5, 500 μl PCI and vortex samples vigorously for 30 s. Spin the samples with 20,000 × *g* for 5 min at room temperature and immediately collect upper aqueous phase (approximate volume 350 μl) to a new 1.5 ml microcentrifuge tube (*see* **Note 35**).
11. Add 1 μl GlycoBlue and 1 ml ethanol. Incubate the samples at −20 °C, overnight (*see* **Note 36**). Pellet the RNA by centrifugation with 20,000 × *g* for 30 min at 4 °C. Discard the supernatant.
12. Wash the pellet twice with 750 μl 70 % ethanol pre-chilled at −20 °C. After each wash spin the pellet for 10 min with the same speed and temperature settings, remove the supernatant, and air-dry the pellet (*see* **Note 37**).

3.5 RNA Cloning (2): 5′ Adapter Ligation in Solution

1. *PNK phosphorylation*. Prepare 15 μl phosphorylation mixture per sample: 11 μl water, 1.5 μl 10 × T4 RNA ligase reaction buffer, 1.5 μl 10 mM ATP (final concentration 1 mM), and 1 μl T4 PNK (final concentration 0.67 U/μl). Resuspend RNA pellet in the reaction mixture and incubate the samples for 30 min at 37 °C.

2. *5′ adapter ligation*. Prepare 5 µl ligation mixture per sample: 2 µl water, 0.5 µl 10× T4 RNA ligase reaction buffer, 0.5 µl 10 mM ATP (final concentration 1 mM), 1 ml 100 µM barcoded 5′ adapter (final concentration 5 µM), and 1 µl T4 RNA ligase 1 (final concentration 0.5 U/µl). Add the ligation mixture to the reaction mixture from **step 1** and incubate the samples for further 6 h at 16 °C (*see* **Note 38**).
3. *RNA purification and concentration*. For easier handling, increase sample volume to 400 µl. Add 50 µl NaOAc, 500 µl PCI and vortex the samples vigorously for 30 s. Spin the samples with 20,000 × *g* for 5 min at room temperature and immediately collect upper aqueous phase (approximate volume 350 µl) to a new 1.5 ml microcentrifuge tube.
4. Add 1 µl GlycoBlue and 1 ml ethanol. Incubate the sample at −20 °C overnight (*see* **Note 39**). Pellet the RNA by centrifugation with 20,000 × *g* for 30 min at 4 °C. Discard supernatant.
5. Wash the pellet twice with 750 µl 70 % ethanol pre-chilled at −20 °C. After each wash spin the pellet for 10 min with the same speed and temperature settings, remove the supernatant, and carefully air-dry the pellet.

3.6 Reverse Transcription and Library Amplification by PCR

1. *Reverse transcription*. Prepare 13 µl RT mix I per sample: 8 µl water, 4 µl 2.5 mM dNTPs (concentration in the final RT reaction mixture 0.5 mM), 1 µl 10 µM miRCat-33 primer (concentration in the final RT reaction mixture 0.5 µM). Resuspend the pellets thoroughly in the RT mixture and incubate the samples for 3 min at 80 °C. Immediately transfer the samples on ice and incubate for 5 min.
2. Prepare 6 µl RT mix II per sample: 4 µl 5× first-strand buffer, 1 µl 0.1 M DTT (concentration in the final RT reaction mixture 5 mM), and 1 µl RNAsin (concentration in the final RT reaction mixture 2 U/µl). Add RT mix II and incubate the samples for 3 min at 50 °C.
3. Add 1 µl SuperScript III Reverse Transcriptase and incubate the samples for 60 min at 50 °C.
4. Inactivate reverse transcriptase by incubation for 15 min at 65 °C. Then degrade template RNA by the addition of 2 µl RNase H and incubation for 30 min at 37 °C (*see* **Note 40**).
5. *PCR amplification of the library*. Prepare 200 µl PCR mixture per sample (*see* **Note 41**): 140 µl water, 20 µl 10× LA buffer, 20 µl 2.5 mM dNTPs (final concentration 0.25 mM), 4 µl 10 µM primer P5 (final concentration 0.2 µM), 4 µl 10 µM primer PE_miRCat_PCR (final concentration 0.2 µM), 10 µl template cDNA from **step 4**, and 2 µl TaKaRa LA Taq (final concentration 0.05 U/µl). For the amplification divide the mixture into four PCR tubes and run the following program: initial

1. Illumina sequencing data

```
@HWI-ST300:149:B09LYABXX:3:1104:20554:107511 1:N:0:
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCGCCGGTTGGAATTCTCGGGGGCCAAAGCCCGGGATGCCGA
+
FAFHCFHFGHFHFEIG>GBG=FF;E;BCFHG@FGGFHIGFI<?BFIGGGECGCDHBA8>A>8/3:5@55;55<&&5<BD###############
```

2. Adapter trimming

```
@HWI-ST300:149:B09LYABXX:3:1104:20554:107511 1:N:0:
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCGCCGGT
+
FAFHCFHFGHFHFEIG>GBG=FF;E;BCFHG@FGGFHIGFI<?BFIGGGECGCDHBA8>A
```

3. Quality filtering and data collapsing

```
>L2HS-344215_4
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCGCCGGT
```

4. Mapping reads to the database (BLAST)

```
L2HS-344215_4  HOXC8_mRNA         100.00  38   0   0   23  60  1990  2027  8e-14  75.8
L2HS-344215_4  miR-196a_miRNA     100.00  22   0   0   1   22  1     22    3e-04  44.1
```

5. Chimera calling

```
L2HS-344215_4 miR-196a_miRNA  1  22  1  22  3e-04 HOXC8_mRNA  23  60  1990 2027 8e-14
```

6. Identification of RNA-RNA interaction sites (UNAFold)

```
344215_4
TAGGTAGTTTCATGTTGTTGGGAAGAAAACCCAACAACTGAGACTGCCTAGCCCGCCGGT
TAGGTAGTTTCATGTTGTTGGG----------------------------------------   miR-196a_miRNA  1     22
--------------------AAGAAAACCCAACAACTGAGACTGCCTAGCCCGCCGGT   HOXC8_mRNA      1990  2027
((((((((((((.(((((((((.......)))))))))))))))))))).........   (-28.5)
```

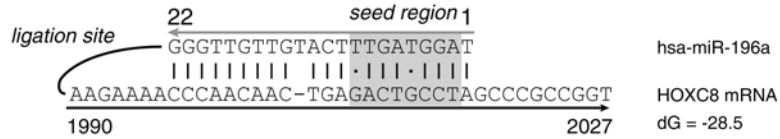

Fig. 3 Outline of the basic bioinformatic analysis of CLASH data. This protocol is incorporated into the hyb pipeline, available from https://github.com/gkudla/hyb and described in detail in [11]

denaturation—2 min at 95 °C; 19–24 cycles of [denaturation—20 s at 98 °C, annealing—20 s at 52 °C, extension—20 s at 68 °C]; final extension—5 min at 72 °C (*see* **Note 42**).

6. Concentrate DNA libraries using MinElute PCR purification kit according to the manufacturer's instructions (*see* **Note 43**). Elute your samples with 20 μl EB buffer provided with the kit.
7. *Adjusting library size distribution*. Prepare 3 % high resolution MetaPhor agarose gel in 1× TBE buffer with SybrSafe according to the manufacturer's instructions (*see* **Note 44**). Wait until it settles and then chill at 4 °C for at least 20 min before run.
8. Place the gel tank on ice for the duration of the run to prevent excessive warming of the agarose gel. Load the samples with added DNA loading dye, a 50 bp size marker and resolve the gel

in 1× TBE buffer for about 2 h at constant voltage (80 V) until bromophenol blue reaches the edge of the gel (*see* **Note 45**).

9. Scan the gel on the gel scanner and print the scan in its original size. Place the gel on the transparency film and carefully align it with the scan. Cut out the DNA fragments of desired length as indicated in Fig. 2b (*see* **Note 46**).
10. Purify DNA library from the gel using MinElute Gel Extraction kit according to the manufacturer's instruction (note on the gel dissolving step). Elute ready DNA libraries with 20 μl EB buffer provided with the kit (*see* **Note 47**).
11. Measure the amount of DNA in the libraries using Qubit fluorimeter and Qubit dsDNA HS Assay Kit.

4 Notes

The CLASH procedure is long with relatively few steps where progress of the experiment can be controlled. It is therefore useful to be observant and note any information that can help with potential troubleshooting. Do: (1) collect western blot samples during the AGO-RNA purification steps to control for the AGO recovery during the procedure and (2) monitor the radioactivity of the sample and all the fractions that are discarded on the way to confirm that the enzymatic steps that can be monitored are performed successfully.

1. For the CLASH procedure expression of tagged AGO protein is essential. We used PTH-AGO1-Flp-InT-REx 293 cells created from Flp-In T-REx 293 cell line (Life Technologies) stably transfected with pcDNA5/FTR/TO vector (Life Technologies) expressing PTH-tagged human AGO1 after Doxycycline induction (described in detail in [10]). PTH tag consists of (1) two immunoglobulin-binding Z domains from *Staphylococcus aureus* Protein A; (2) TEV protease cleavage site (not exploited in this protocol); and (3) 6×His-tag. However, we believe that the only part that is essential for generating reliable CLASH data is the 6×His tag, which allows for the very stringent, denaturing purification of AGO-RNPs, thus contributing to the low background of the method [10]. The first step of purification can probably be performed with any other good-efficiency tag in place of Protein A. In other analyses we have successfully used a tripartite tag with FLAG—PreScission protease cleavage site—6×His for RNA-protein cross-linking [15]. The unmodified parental cell line served as negative control. Stably transfected cell line can potentially be replaced with transiently transfected cells.

2. Enzymes provided by the given suppliers were tested in our lab and give consistently low background. If another source is to be used, check first with the supplier that the enzyme is not His-tagged (information is not always included in the enzyme datasheet). Be aware that commercial enzymes are frequently contaminated with varying amounts of RNA originating from the organism that was used for the enzyme production. Although this free RNA can usually be removed during Ni-WBI washes on Ni-NTA beads, foreign RNA can especially affect samples during enzymatic reactions in solution.
3. RNase treatment is a very important step in the CLASH protocol. The optimal length of the RNA fragments is around 25 nucleotides; that is, they are long enough to be reliably mapped to the database after cloning and sequencing and not too long for the RNA-RNA ligation (non-basepaired loop regions between ligated RNA fragments are usually shorter than 10 nucleotides [10]). Therefore RNase treatment is the step that should be optimized during the first experiment. Commercially available RNase stock is highly concentrated, so to minimize discrepancies between experiments it is practical to prepare a working stock of RNases (we use 1:20 dilution in water), store it at 4 °C, and use for all subsequent experiments.
4. Our custom designed 5′ adapters have the following structure: (1) 5′ blocking group to prevent self-ligation, (2) common sequence required for the Illumina sequencing platform, (3) random barcode (3 nucleotides) that allows for identification of PCR duplicates during bioinformatic analysis of sequencing data, and (4) sample-specific 5–8-nucleotide barcode that allows for samples multiplexing for Illumina sequencing. These adapters and matching PCR primers can potentially be replaced by many others used successfully in similar experiments. To prevent sample cross-contamination, use fresh aliquots of adapters for each experiment.
5. Wet-transfer apparatus allows for transfer of the AGO-RNA complexes without excessive heating, which could result in RNA degradation.
6. IgG coated Dynabeads should be prepared in advance according to the protocol in [14]. Alternatively IgG Sepharose 6 Fast Flow (GE Healthcare) can probably be used. We have successfully used it for experiments other than AGO protein-RNA complexes.
7. If expression of the tagged AGO protein is inducible, include an induction step in the protocol. Control for the expression level of tagged AGO protein in the cells, since overproduction can lead to the generation of spurious binding sites (note that Protein A binds nonspecifically to almost all antibodies and for

reliable comparison of endogenous and exogenous AGO levels it should be removed by TEV cleavage).

8. To minimize cells exposure to changing conditions, reduce the time between taking cells out of the incubator and crosslinking by proceeding with one dish at a time.
9. Cell lysates can be used directly, or aliquots can be flash frozen in liquid nitrogen and stored at −80 °C for future use.
10. Western blotting is not essential for CLASH analysis but it is a good idea to monitor the course of the experiment by preparing samples at points during the CLASH protocol. This allows potential problems with AGO-RNA purification steps to be identified.
11. It is important to wash the IgG-Dynabeads well before combining with the lysates. Although the coating of the magnetic beads is covalent, IgGs can be released into the solution, affecting the purification procedure by competing with the resin for the tagged AGO binding.
12. To minimize variation between samples, sufficient enzymatic reaction mix should be prepared in batch for all samples, with an additional 10 % volume for pipetting errors.
13. "Ni-NTA input" and "Ni-NTA flow through" samples cannot be prepared by simple boiling with loading buffer as they contain high concentration of GuHCl. It can be removed by TCA precipitation and extensive acetone washes.
14. Sepharose beads can clog some pipette tips. To prevent this, trim tips at the end with a clean scalpel.
15. At this stage all the western blot samples ("IgG input" Subheading 3.1, **step 4**; "IgG flow through" Subheading 3.2, **step 2**; "IgG beads" Subheading 3.2, **step 5**; "Ni-NTA input" Subheading 3.2, **step 5**; "Ni-NTA flow through" Subheading 3.2, **step 7**) are ready. For AGO detection (protein A tagged) very sensitive peroxidase-anti-peroxidase soluble complex antibody (PAP; Sigma-Aldrich) can be used. We routinely observe significant loss of AGO-RNA complexes during the procedure, ~30 % and 25 % loss at IgG Dynabeads and Ni-NTA purification steps, respectively.
16. Extensive washing with Ni-WBI helps remove non-crosslinked RNA from the CLASH sample. That is very important for minimizing the recovery of background RNA and obtaining high quality CLASH libraries.
17. A metal rack for 1.5 ml microcentrifuge tubes greatly simplifies working with snap cap columns. It helps keep columns vertical and cold when placed in the ice bucket.
18. All washes are performed under gravity flow. The same wash collection tubes can be reused multiple times. GuHCl present

in the samples loaded on the columns affects protein stability and therefore must be removed completely before subsequent enzymatic reactions. Remember to rinse the column side walls especially well during the PNK-WB wash that directly precedes the enzymatic steps.

19. For convenience follow this order: (a) close the lid to remove the remains of the buffer from previous step; (b) close the snap cap and place the column in the collection tube; (c) open the lid and add the enzymatic reaction mixture; (d) close the lid.
20. It is very important to open the columns in the following order: (a) open the column lid; (b) open the snap cap and quickly transfer columns to their collection tubes. This prevents quick dripping of the buffer from the column and thus potential cross-contamination of the samples or, in further steps, radioactive contamination of the workspace. For subsequent more efficient rinsing of GuHCl traces from the column carefully add the Ni-WBI buffer directly to the center of the column.
21. At this stage the following RNA ends should be present in the sample: 5′ P after T4 PNK treatment, 3′ P created directly by the RNase cleavage and 3′OH of the full-length miRNAs (miRNAs are short and have high chance of escaping RNase digestion) and potentially other non-cleaved RNA species bound by AGO. As T4 RNA ligase requires 3′ OH, the ligation products should be enriched in full-length miRNAs, and miRNA targets should be favored in the reaction because of their proximity to miRNAs. Intermolecular RNA-RNA ligation is the last enzymatic step that is performed at low temperature. After this step low-temperature stabilization of RNA-RNA interactions is not needed and should not affect the CLASH result.
22. The labeling reaction is performed in the presence of radioactive ATP only. This concentration is enough to obtain a reasonable radioactive signal from the RNA in the sample but is too low for efficient phosphorylation of all RNA 5′ ends in the sample. Therefore, to convert all the RNA 5′ ends before 5′ end adapter ligation, an additional phosphorylation reaction is performed at Subheading 3.5, **step 1**.
23. Removal of most free radioactive ATP decreases the chance of radioactive contamination at later stages. Perform the washes until the radioactivity of the flow through measured with a manual radioactivity monitor falls to approximately 10 cps.
24. The approximate volume of eluate is 1 ml and is suitable for further TCA precipitation. However if acetone precipitation is to be performed in the following steps (*see* **Note 25**), the volume of the eluate must be reduced. Perform two rounds of elution, each with 125 μl elution buffer. The radioactive signal

of the eluate, although occasionally misleading, usually gives good estimation of the amount of AGO-RNPs present in the sample. Higher radioactivity in experimental samples compared to negative controls is expected. This difference should be even more pronounced after TCA precipitation.

25. The supernatant after TCA precipitation still contains significant amounts of unincorporated ^{32}P-γ-ATP, so high radioactivity is expected. The AGO-RNP pellet that is formed after TCA precipitation can occasionally be difficult to resuspend. If this is a problem, TCA precipitation can be replaced with acetone precipitation. To perform this, modify the elution step as described in the previous note, then mix 250 μl of the eluate with 1.25 ml acetone and incubate samples overnight at –20 °C. Continue with the centrifugation and washes (a single acetone wash is enough) as described for TCA precipitation.
26. Pellets formed by AGO-RNPs should be small and clear. If it is big and white, it contains GuHCl carried over from Ni-WBI buffer that has coprecipitated with the RNPs. Wash the samples really well with acetone and, if needed, add additional acetone wash step. Longer incubation with acetone on ice can also be considered.
27. Alternatively, a thermoblock set to 37 °C can be used. However, it must be done with great care since over-dried pellets may be difficult or even impossible to resuspend.
28. After addition of NuPAGE LDS sample buffer observe the color of the samples. A greenish tint indicates residual TCA in the sample that can lead to uneven migration during electrophoresis. In such a case add ~1 μl 1 M Tris–HCl pH 7.8 and the color of the sample should become blue. To prevent RNA degradation during sample heating, do not increase the temperature or incubation time at 65 °C. Prepared samples can be stored at –20 °C for a few days. However, the radioactive signal will gradually decrease, so the storage time should not be unnecessarily extended.
29. At this stage the samples are not yet barcoded, so minimize the chance of cross-contamination by leaving a single-lane space between samples. NuPAGE MOPS SDS buffer that is used for the gel electrophoresis in CLASH is designed to maintain similar pH during a whole gel separation process. Do not replace it with the Tris-based SDS running buffer, in which the pH significantly increases during electrophoresis potentially causing RNA degradation. Although NuPAGE gels can be run at voltages higher than 100 V, avoid it as it can lead to significant gel heating and also result in RNA degradation.
30. The wet-transfer system is used for its low tendency to overheat. In addition we usually cool the transfer tank during the

run by putting stirring bar inside the buffer chamber and placing the tank in an ice box on the stirring plate.

31. A phosphorescent ruler (or even two at the same time) attached to the cling film is crucial for the accurate alignment of the nitrocellulose membrane with the autoradiography film. The time required for membrane exposure to the film can vary significantly from experiment to experiment. However, with the recommended amount of starting material and a successful procedure, 1 h exposure should give a suitable signal. Low signals suggest problems with HTP-AGO expression, UV cross-linking, the RNP purification procedure (which can be controlled for by the western blot with samples collected during IgG Dynabeads and Ni-NTA purification), radioactive RNA labeling reaction, or contamination with RNases.
32. A typical autography film is shown in Fig. 2a. In the experimental samples a major, broad radioactive signal around 100 kDa is expected, corresponding to the AGO-RNA complexes. A weaker signal, possibly originating from aggregated complexes, is frequently also visible in the 200 kDa range. The negative control should not produce much radioactive signal and after prolonged (e.g., overnight) exposure a uniform signal from the whole lane is expected. A strong background signal in experimental samples and negative controls indicates substantial unspecific RNA binding and correlates with low quality results. In this case, try extending the Ni-WBI washes during Ni-NTA purification.
33. To cut out the membrane fragments mark the film with boxes around the radioactive bands to be excised, align the film with the membrane and make punctures at the corners of the boxes through the film and a membrane with an injection needle. Remove the film. Marks left by the needle make it easy to excise membrane fragments with a scalpel. Exposed membrane or cut out membrane fragments may be stored at −20 °C for longer periods, since the radioactivity of the sample is no longer crucial.
34. It is useful to measure the approximate radioactivity of the membrane fragments (using a regular radioactivity monitor) before the Proteinase K treatment, and the same fragments and Proteinase K solution after the procedure. The signal from the membrane should be significantly decreased after treatment. This additional control gives an indication of the efficiency of the Proteinase K treatment. While adding more Proteinase K or longer incubation of the samples could potentially solve the problem, our experience suggests that RNA retention on the nitrocellulose membrane is a result of the preceding steps and cannot be readily resolved at this stage.

35. During aqueous phase collection, be very careful not to disturb the interphase or aspirate the lower, organic phase. When not processed quickly after the centrifugation, samples tend to lose clarity. In such a case repeat the centrifugation of unprocessed samples. Afterwards compare the radioactivity of the collected aqueous phase with the organic phase. Radioactive signal in the organic phase suggests inefficient Proteinase K treatment leading to the protein fragments remaining bound to the RNA molecules, preventing them from entering aqueous phase.
36. Alternatively ethanol precipitation can be performed at −80 °C for 30 min. Avoid longer incubation as it can result in adverse salt co-precipitation. The RNA pellet after ethanol precipitation and centrifugation should be attached to the microcentrifuge tube walls. Loose pellets may indicate the presence of organic phase traces after PCI extraction. RNA pellets can be stored indefinitely at −20 °C.
37. Alternatively pellets can be dried at 37 °C with care not to over-dry them. Over-dried pellets are difficult or impossible to resuspend.
38. According to the enzyme manufacturer, the presence of T4 PNK should not affect the ligation reaction. This reaction can also be performed overnight.
39. RNA pellets after ethanol precipitation can be stored indefinitely at −20 °C.
40. Prepared cDNA libraries can be stored long term at −20 °C.
41. It may be informative to first perform a small-scale amplification using 1 μl cDNA as an indication of the amount of material obtained and the quality of the library. The quality of the library can be assessed further by PCR amplification with miRNA-specific primers, e.g., miR-16 that is abundant in many tissues and cell lines (PCR-1 primers: miR-16F: GCAGCACGTAAATATTGGCG and PE_miRCat_PCR; PCR-2 primers: miR-16R: GCCAATATTTACGTGCTGCTA and P5). The presence of the abundant miRNA-specific bands confirms efficient protein-RNA cross-linking and adapter ligation. To increase the sequencing depth of the DNA library for high-throughput sequencing—which is especially crucial for chimeric sequences—perform two independent PCR amplifications and use the entire amount of cDNA template.
42. When preparing the DNA library for high-throughput sequencing, limit the number of amplification cycles to minimize the frequency of PCR duplicates. Adjust the number of cycles to obtain ~20–80 ng DNA from the total PCR.
43. The wash buffer provided with the PCR purification kit contains ethanol, which should be removed completely by, e.g.,

air-drying columns for 5 min just before elution. Remaining ethanol traces can lead to the samples diffusing from wells during loading. If this happens, save the remaining samples by incubation in open tubes for a few minutes at 50 °C, and then try loading again.

44. As preparing Metaphor gels is time consuming, it is convenient to prepare the gel the previous day, wrap in cling film, and store chilled at 4 °C.
45. At this stage the samples are already barcoded and cross-contamination of samples should not be a problem; however, it is still a good idea to load the samples leaving a lane space in between.
46. A typical picture of the DNA library resolved on an agarose gel is presented at Fig. 2b. A lower band of about 120 bp is frequently visible and corresponds to the amplified sequencing adapter dimers. Above it, starting at about 140 bp, a wide, smeary band is visible. The bottom of the smeared band contains most DNA material is sharper, and consists largely of cloned miRNAs. The region above is enriched for miRNA targets and chimeric reads. The presence of a sharp band may indicate excessive RNA digestion. Lack, or small amounts, of PCR products on the agarose gel, despite strong signal by autoradiography, suggests inefficient enzymatic reactions due to problems with the enzymes or sequencing adapters. At this stage it is possible to adjust library size distribution and enrich the DNA library for chimeric sequences. For this, isolate DNA library from the gel in two fractions—lower (LB) ~140–150 bp and upper (UB) ~150–210 bp as indicated on the figure. Before sequencing combine both fractions in weight ratio LB:UP 1:3.
47. Using TaKaRa polymerase for the PCR amplification allows TOPO-TA cloning of the DNA library and subsequent small scale Sanger sequencing (e.g., Life Technologies kit; use 2 μl DNA library per 6 μl cloning reaction). This should provide information on (1) RNA biotypes in the original sample, expect high proportion of miRNAs; (2) length of DNA fragments in the library; and (3) potential contaminations which cannot be deduced from the DNA library resolution on the MetaPhor gel.

References

1. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505:344–352
2. Bernstein BE, Birney E, Dunham I et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
3. Peterson SM, Thompson JA, Ufkin ML et al (2014) Common features of microRNA target prediction tools. Front Genet 5:23
4. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233

5. Ha I, Wightman B, Ruvkun G (1996) A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. Genes Dev 10:3041–3050
6. Lal A, Navarro F, Maher CA et al (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3′UTR microRNA recognition elements. Mol Cell 35:610–625
7. Kedde M, van Kouwenhove M, Zwart W et al (2010) A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility. Nat Cell Biol 12:1014–1020
8. Ule J, Jensen KB, Ruggiu M et al (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215
9. Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci U S A 106:9613–9618
10. Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654–665
11. Travis AJ, Moody J, Helwak A et al (2013) hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. Methods 65:263. doi:10.1016/j.ymeth.2013.10.015
12. Helwak A, Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nat Protoc 9:711–728
13. Grosswendt S, Filipchyk A, Manzano M et al (2014) Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol Cell 54:1042–1054
14. Oeffinger M, Wei KE, Rogers R et al (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. Nat Methods 4:951–956
15. Tree JJ, Granneman S, McAteer SP et al (2014) Identification of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. Mol Cell 55:199–213

Part V

RNA Modifications

Chapter 15

Genome-Wide Analysis of A-to-I RNA Editing

Yiannis A. Savva, Georges St. Laurent, and Robert A. Reenan

Abstract

Adenosine (A)-to-inosine (I) RNA editing is a fundamental posttranscriptional modification that ensures the deamination of A-to-I in double-stranded (ds) RNA molecules. Intriguingly, the A-to-I RNA editing system is particularly active in the nervous system of higher eukaryotes, altering a plethora of noncoding and coding sequences. Abnormal RNA editing is highly associated with many neurological phenotypes and neurodevelopmental disorders. However, the molecular mechanisms underlying RNA editing-mediated pathogenesis still remain enigmatic and have attracted increasing attention from researchers. Over the last decade, methods available to perform genome-wide transcriptome analysis, have evolved rapidly. Within the RNA editing field researchers have adopted next-generation sequencing technologies to identify RNA-editing sites within genomes and to elucidate the underlying process. However, technical challenges associated with editing site discovery have hindered efforts to uncover comprehensive editing site datasets, resulting in the general perception that the collections of annotated editing sites represent only a small minority of the total number of sites in a given organism, tissue, or cell type of interest. Additionally to doubts about sensitivity, existing RNA-editing site lists often contain high percentages of false positives, leading to uncertainty about their validity and usefulness in downstream studies. An accurate investigation of A-to-I editing requires properly validated datasets of editing sites with demonstrated and transparent levels of sensitivity and specificity. Here, we describe a high signal-to-noise method for RNA-editing site detection using single-molecule sequencing (SMS). With this method, authentic RNA-editing sites may be differentiated from artifacts. Machine learning approaches provide a procedure to improve upon and experimentally validate sequencing outcomes through use of computationally predicted, iterative feedback loops. Subsequent use of extensive Sanger sequencing validations can generate accurate editing site lists. This approach has broad application and accurate genome-wide editing analysis of various tissues from clinical specimens or various experimental organisms is now a possibility.

Key words *Drosophila melanogaster*, RNA editing, ADAR, Double-stranded RNA, Transcriptome, Protein recoding, Noncoding RNAs, Neurological disorders, Next-generation sequencing, Single-molecule sequencing, Inosinome

1 Introduction

1.1 A-to-I RNA Editing

Mature RNA molecules often vary substantially from their genomic origins via posttranscriptional RNA processing events such as alternative splicing. However, more subtle changes in mature RNAs can occur through RNA editing [1]. The most

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_15,

prevalent and evolutionarily conserved RNA-editing system is the deamination of adenosine-to-inosine (A-to-I). This phenomenon involves the conversion of adenosine nucleotides into inosine through hydrolytic deamination (Fig. 1a) mediated by *a*denosine *de**a*minases acting on *R*NA (ADAR) [2]. RNA editing enzymes consist of double-stranded RNA-binding domains (dsRBDs) as well as a catalytic domain in the C-terminal part of the protein [3]. ADAR targets duplex RNAs of various structural arrangements and lengths. The structural variability of RNA substrates confers two distinct types of editing specificities. For example, short imperfect dsRNA molecules containing mismatches, bulges, and loops are edited specifically while long perfectly base-paired dsRNAs are edited promiscuously [2]. Inosine nucleosides mimic the base pairing properties of guanosine through the formation of Watson-Crick bonds with cytosine (Fig. 1a). Therefore, the cellular machinery interprets inosines as guanosines [4]. Specific RNA editing in coding regions has the capacity to recode the genome via amino acid substitutions in highly conserved and functionally important residues within proteins [5]. For example, the rate of inactivation in potassium channels is regulated by specific RNA-editing events, that result in non-synonymous amino acid reassignments [6]. The A-to-I RNA editing system is highly active in the nervous system and edits transcripts encoding products which are involved in electrical and chemical neurotransmission, such as components of the synaptic release machinery as well as ligand-gated and voltage-gated ion channels [7]. Specific editing is also active in non-coding sequences and is associated with regulating RNA splicing through creation or elimination of splicing signals and with the regulation of biogenesis and function of microRNAs (miRNAs) [2]. Promiscuous ADAR editing activity occurs invariably in noncoding regions of the genome and this activity is typically observed in transposable element sequences embedded in introns, within untranslated regions (UTRs), and in long noncoding RNAs (lncRNAs) [8, 9]. This form of abundant editing is involved in nuclear retention of transcripts [10], in cellular defense against viral RNAs [11], and in regulation of RNA interference (RNAi) pathways [3]. Phenotypes of ADAR deficiencies in various model organisms provide evidence that appropriate nervous system function requires an adequate A-to-I RNA-editing activity. Loss of RNA editing in invertebrates results in severe neurological defects and diverse behavior abnormalities. In *C. elegans*, loss of RNA editing results in chemotaxis defects [12]. Furthermore, *Drosophila* editing mutants exhibit coordination defects, seizures, temperature-sensitive paralysis, defects in courtship display, and age-dependent neurodegeneration [13]. More severely, the loss of editing activity in mammals leads to lethality. Specifically, the deletion of ADAR1 editing enzyme is embryonic lethal due to hematopoiesis defects and elevated cellular apoptosis

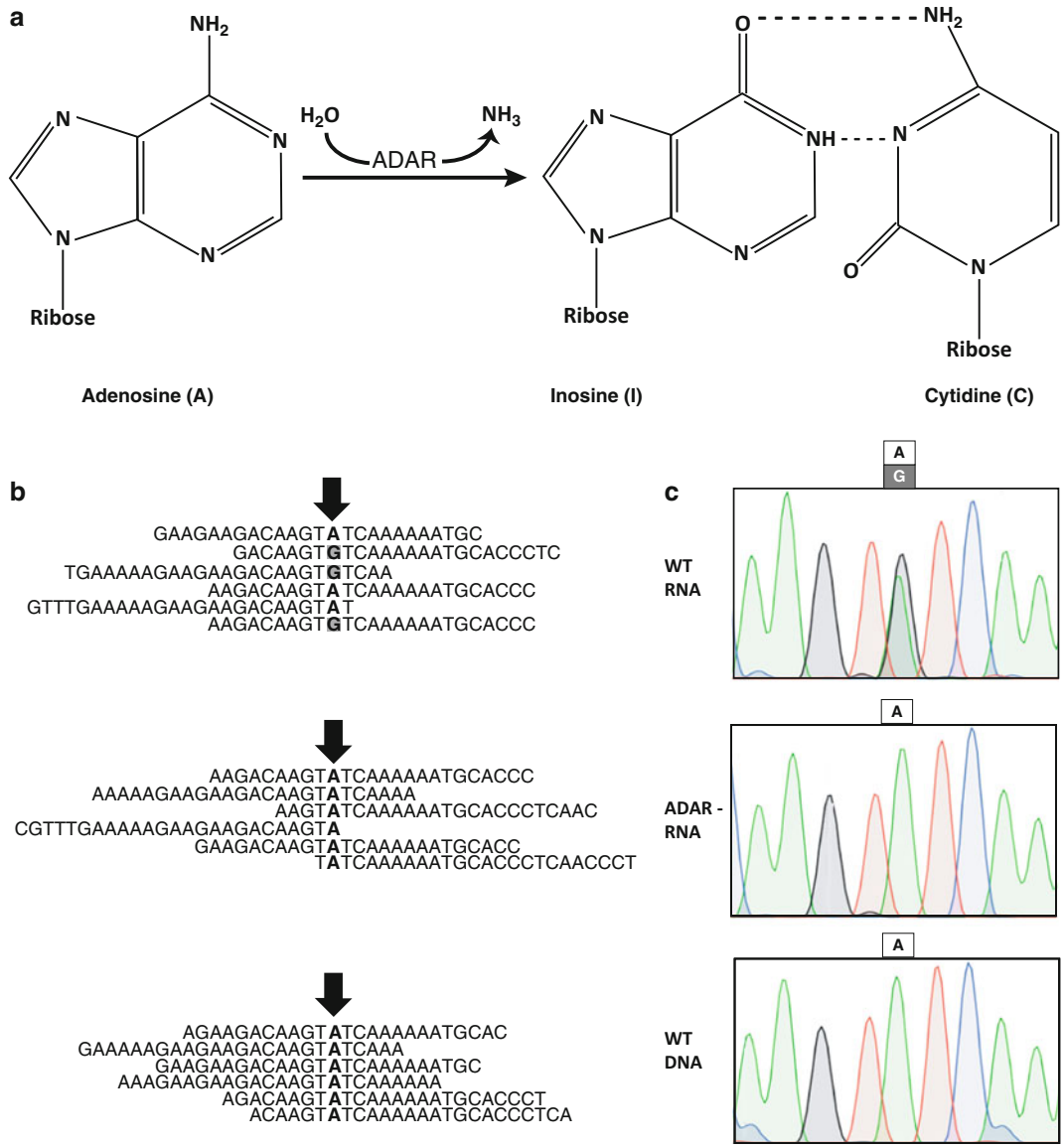

Fig. 1 The hydrolytic deamination of adenosine to inosine. (**a**) An adenosine is converted to inosine via the hydrolytic deamination of an adenine base. Inosine shares the binding properties of guanosine, and thus forms bonds with cytidine. (**b**) Sequences generated through deep sequencing technologies contain a mixture of edited (G) and unedited (A) reads after proper alignments. Contrary, sequences generated from ADAR-deficient and wild-type DNA samples invariably contain unedited (A) reads. (**c**) The signature of A-to-I RNA editing. Example electropherograms generated by Sanger sequencing of cDNA molecules exhibit mixed A/G peaks in wild-type RNA sample. In contrast, ADAR-deficient and wild-type DNA electropherograms contain only the genomic encoded version in sequences

[14]. Similarly, ADAR2 deletion also results in postnatal lethality caused by severe seizure episodes [15]. These phenotypes highlight an important role for the posttranscriptional process of A-to-I RNA editing in metazoan physiology.

1.2 A-to-I RNA Editing and Neurological Disorders

Analysis of transcriptional landscapes within 15 human cell lines via deep sequencing technologies revealed that the majority of the genome is transcribed [16]. Indeed, pervasive transcription produces vast numbers of RNA molecules originating from noncoding portions of the genome [17]. These lncRNAs participate in diverse cellular functions during mammalian development, such as dosage compensation, genomic imprinting, and cell differentiation [18], through the formation of intricate secondary and tertiary structures that act as gene regulatory elements [19]. Not surprisingly, surveys of RNA folding in various eukaryotic genomes suggest that the transcriptome occupies a highly complex structural configuration [20], providing an additional informational layer analogous to the genetic code [21]. Moreover, these discoveries suggest that proper cellular function depends on the accurate expression of noncoding and coding RNA molecules, whose orchestrated processing allows for functional specificities. Thus, it is not surprising that mutations within proteins involved in almost all aspects of RNA metabolism lead to cellular catastrophes and various human diseases [22]. In particular, several studies have linked abnormal RNA editing with various neurological disorders [23]. For example, aberrant RNA editing of glutamate receptor is strongly linked to Amyotrophic lateral sclerosis and epilepsy. Likewise, serotonin receptor editing has been implicated in depression, schizophrenia and Prader-Willi syndrome. Alterations in ADAR expression occur in glioblastoma, a brain-specific cancer, and misregulation of dsRNA metabolism mediated by the ADAR1 editing enzyme is linked to Aicardi-Goutieres syndrome, a neurodevelopmental disorder [24].

Despite these links between RNA editing and cellular disease, appropriate tools remain largely unavailable for the measurement of changes in editing levels. Additionally, relatively little data exists on stress and activity induced changes in editing levels. Current research in the field implements next generation sequencing in order to identify authentic RNA-editing sites across a broad range of phyla. Understanding the dynamics of ADAR-mediated editing in a variety of environmental and physiological contexts represents an important avenue of current research.

1.3 Next-Generation Sequencing to Study RNA Editing in Drosophila

Since inosine forms base pairs with cytosine, the cellular machinery recognize inosine as guanosine. This is observed as A-to-G substitutions in RNAseq reads obtained from deep sequencing platforms (Fig. 1b). More importantly, these A-to-G substitutions in reads can be validated using Sanger sequencing. Electropherograms generated from cDNA libraries exhibit a mixed A/G peak at the edited adenosine (Fig. 1c). During the last decade several studies attempted to identify the exact genomic locations of RNA editing sites in various cell lines, tissues and model systems. Although identification of A-to-G substitutions in RNAseq experiments sounds simple, an inherent number of technical and biological

factors can lead to variations and errors in the detection and measurement of RNA-editing sites. For example, a recent study reported that RNA-editing events are mechanistically more widespread than previously thought, leading to all possible nucleotide substitutions in human B cells, thereby expanding the range of RNA editing types [25]. Yet, most of these RNA editing events were attributed to artifacts generated from common sources of errors by next generation sequencing technologies [26–28]. Additionally, editing site discovery studies in *Drosophila* showed relatively poor overlap between sites and uncovered a large number of editing sites that are specific to individual wild type lab stocks [29–31]. Although different lab stocks may carry specific RNA editing events, stock-specific RNA editing alone cannot explain the unprecedented variation observed in these studies (Fig. 2a).

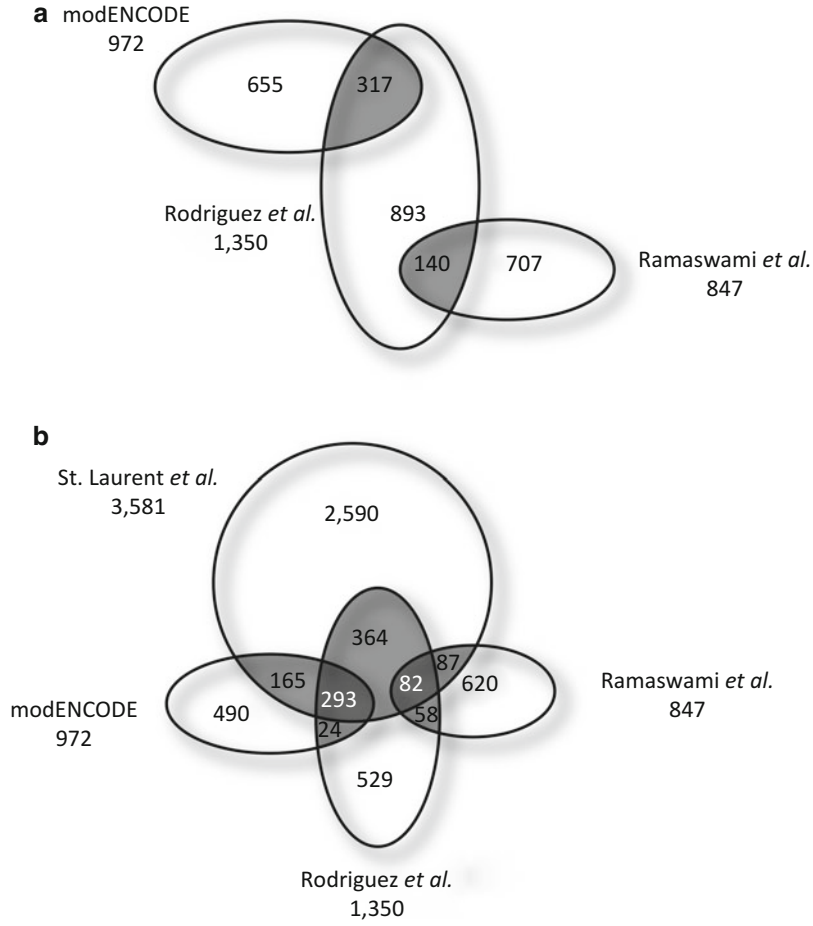

Fig. 2 Comparison of RNA-editing site datasets from recent studies in *Drosophila*. (**a**). Venn diagram showing the relations between three independent *Drosophila* editing datasets identified by next-generation sequencing technologies. RNA-editing sites reported by modENCODE, Rodriguez et al., and Ramaswami et al. exhibit relatively poor overlap. (**b**) Venn diagram showing the relations between RNA-editing sites reported through the method described here (St. Laurent et al.), compared to the other three published RNA-editing lists

Table 1

***Drosophila* A-to-I RNA-editing site discovery pipelines and metrics**

Publication	Year	Novel sites	Sanger validation	% Validation	False positive rate (%)	False negative rate (%)
St. Laurent et al.	2013	3581	1072	29.9	13.6	55.2
modENCODE	2011	877	0	0	58.5	89
Rodriguez et al.	2012	1350	0	0	46.32	83.1
Ramaswami et al.	2012	847	11	1.3	59.39	89.4

To different degrees, the discrepancies seen between the three independent *Drosophila* datasets serve as an important reminder of the quality-control issues that may arise from the use of high throughput next generation sequencing experiments. Additionally, they highlight the need for more rigorous methodologies for the assessment of experimental reproducibility [32].

Here we describe our editing site discovery pipeline protocol, in *Drosophila*, using single-molecule sequencing. This method identified 3581 editing sites (Fig. 2b) and achieved a measure of success in both specificity (false-positive rate) and sensitivity (False Negative Rate) [33]. Coupled with extensive Sanger sequencing validations, the method generated the most accurate editing site discovery dataset in *Drosophila* to date. Our method provides a benchmark necessary to observe meaningful biological patterns resulting from the process of A-to-I RNA editing. Most available datasets do not include sufficient information or validation experiments for quality metrics, such as sensitivity, and specificity. Comparison of our results with other recently published *Drosophila* datasets (Table 1) demonstrates the effectiveness of the protocol. Our pipeline achieves three goals: (a) the identification of putative RNA-editing sites with high validation rate, (b) the successful capture of the majority of editing sites in any given experimental sample, and (c) the usage of this dataset to increase the visibility of ADAR-mediated editing in the context of transcriptome systems biology.

2 Materials

2.1 DNA Preparation

1. Maxwell® 16 Tissue DNA Purification Kit (Promega).

2.2 RNA Preparation

1. TRIzol® reagent (Invitrogen).
2. TurboDNase Buffer (Applied Biosystems).
3. RNaseOut (Invitrogen).
4. TurboDNase (Applied Biosystems).
5. RNeasy MinElute kit (Qiagen).

2.3 Ribosomal RNA Depletion

1. Oligos complementary to the Drosophila 18S and 28S rRNA.
2. DEPC water.
3. 10 mM ddNTP mixture (Roche).
4. 2.5 mM $CoCl_2$.
5. 10× TdT buffer (NEB).
6. Terminal Transferase (NEB).
7. Performa DTR cartridges (EdgeBio).
8. RiboMinus Eukaryote Kit for RNA-Seq (Invitrogen).

2.4 Synthesis of cDNA

1. Superscript III kit (Invitrogen).
2. RNAseIf (NEB).
3. Performa Gel Filtration Columns (EdgeBio).

2.5 PolyA Tailing and 3′ Blocking

1. PolyA Control Oligo (Helicos).
2. 2.5 mM $CoCl_2$.
3. 10× TdT buffer (NEB).
4. PolyA tailing dATP (Helicos).
5. Biotinylated ddATP (Perkin Elmer).
6. USER enzyme (NEB).
7. DEPC water.
8. AMPure beads (Beckman Coulter).
9. 70 % Ethanol.
10. TE buffer.

2.6 Sequencing

1. Helicos Single Molecular Sequencer.

2.7 Validation

1. PCR primers.
2. Sequencing primers.
3. Phusion High-Fidelity PCR Kit (NEB).
4. ExoSAP-IT (Affymetrix).

3 Methods

Highly complex genomic datasets produced with the implementation of next-generation sequencing technologies have relatively high error rates and present challenges in distinguishing patterns of biological knowledge from sources of noise and variation. These sources include unannotated SNPs, alternative splicing events, sequencing platform errors, sequence read misalignments [34], and potentially non-ADAR-mediated RNA sequence alterations [25]. Recently published RNA editing lists were shown to contain numerous false positives [33], likely because minimal Sanger

validation was performed. Similarly, several other recent editing site datasets provided only sparse validation through Sanger sequencing [29, 35, 36]. In order to generate authentic information of sufficient quality to measure both the sensitivity and specificity of RNA editing sites in a whole organism we chose the Helicos single-molecule sequencing platform due to its advantages in transcript detection and reproducibility when compared to other sequencing platforms. Other advantages include unbiased coverage of rarely expressed transcripts [37], minimal sample preparation and avoidance of PCR amplification and ligation [38], and a very low A-to-G substitution error rate [33]. As a model organism, *Drosophila* bestows an ideal system for RNA editing profiling for the following reasons: Firstly, the presence of a well characterized collection of known editing sites exists for *Drosophila* [5]. Secondly, the existence of an ADAR deficiency model [13], and the availability of 15 sequenced genomes from various *Drosophila* species [39] further provide a well-established platform for the study of RNA editing. The method described here couples the depth of SMS with the accuracy of Sanger sequencing to determine *bona fide* RNA-editing events at the genome-wide scale. Specifically, our method uses a three-way comparison between the transcriptomes of wild-type (WT) and ADAR-deficient *Drosophila* and with the resequencing of our WT lab stock genome to comprehensively uncover the inosinome of an adult metazoan organism (Fig. 3).

3.1 DNA Preparation

1. Isolate DNA from whole male Canton-S (wild type) and ADAR null flies using the Maxwell® 16 Tissue DNA Purification Kit following the manufacturer's protocol. Place 10–20 flies of the same genotype in well #1 of the DNA cartridge. Place a plunger in well #7 of the DNA cartridge. Add 500 μl of the elution buffer to the elution tube. Place the DNA cartridge and elution tube in a Maxwell 16 robot to isolate genomic DNA from *Drosophila* tissue.

3.2 RNA Preparation

1. Isolate total RNA from whole male Canton-S (wild type) flies using TRIzol® reagent (Invitrogen) following the manufacturer's protocol.
2. Mix 20 μg of total RNA with 10 μl TurboDNase Buffer; 1 μl RNaseOut; and 2 μl TurboDNase. Incubate the reaction for 30 min at 37 °C.
3. Purify the RNA using the RNeasy MinElute Kit following the manufacturer's protocol.

3.3 Ribosomal RNA Depletion

1. Deplete rRNA from total RNA samples using the RiboMinus Eukaryote Kit for RNA-Seq following the manufacturer's protocol (*see* **Note 1**).

3.4 Synthesis of cDNA

1. Use between 100 and 200 ng of rRNA-depleted RNA for cDNA synthesis using the Superscript III Kit (*see* **Note 2**).

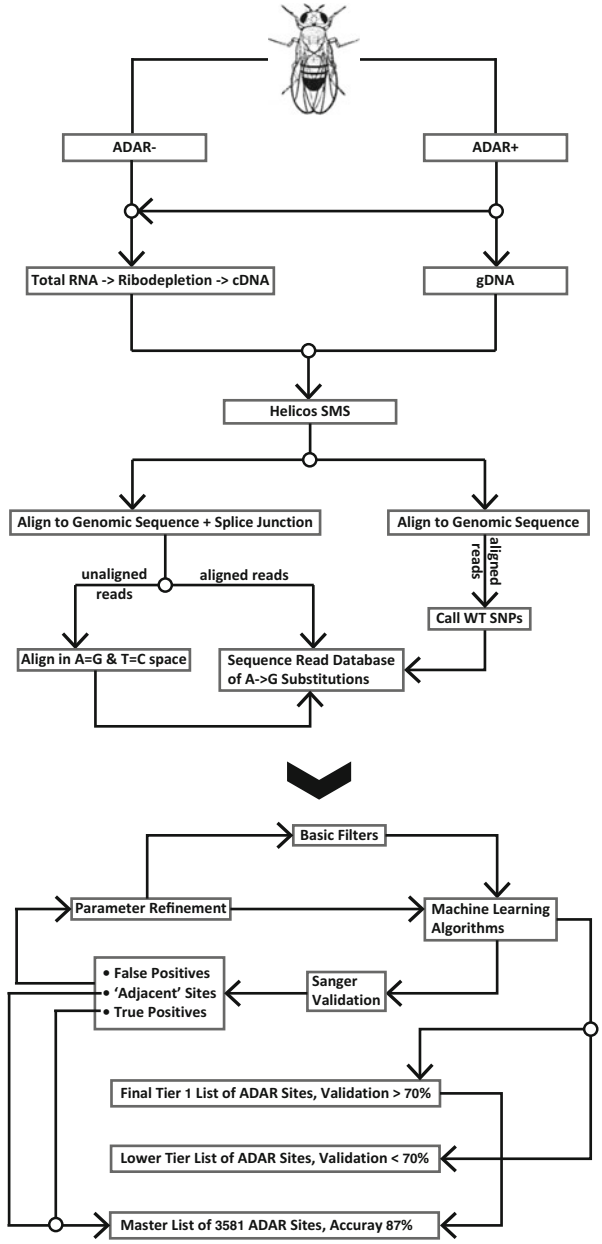

Fig. 3 A schematic diagram of the discovery pipeline for the *Drosophila* inosinome. Total (rRNA-depleted) RNA from adult wild-type flies served as the starting material for the discovery of novel A-to-I RNA-editing events. To account for single-nucleotide polymorphisms (SNPs) and various other artifacts not related to the RNA-editing process, we additionally sequenced DNA from wild-type flies and RNA from ADAR-deficient samples. Single-molecule sequencing reads were examined for the highest alignment quality, using an alignment pipeline designed to select the highest quality alignments while completely avoiding penalties for A-to-G substitutions, the signature of A-to-I RNA editing. Subsequent analysis generated a database of possible novel RNA-editing sites. With the implementation of Basic Filters and Machine Learning Algorithms the editing database was filtered further to distinguish between real editing sites and the many different kinds of false positives. One of the key features of this discovery pipeline is a strong reliance on validation of randomly selected sites through Sanger sequencing to generate True Positives and True Negative sites, and further train the Machine Learning Algorithms at each iteration. Using this repetitive computational approach the final version of the Machine Learning Algorithm was used to partition possible RNA editing sites to establish the Tier 1 list (conservative thresholds), and the Tier 2 list (medium thresholds). Finally, with the implementation of Sanger sequencing the validation rates for the two Tier lists were confirmed by sequencing random selected sites

3.5 PolyA Tailing and 3′ Blocking

1. Use 100 ng of cDNA in 28 μl water and add 5 μl of Helicos PolyA Control Oligo. Incubate the reaction for 5 min at 95 °C and rapid cool on ice.
2. Add 5 μl 2.5 mM of $CoCl_2$, 5 μl of Helicos PolyA tailing dATP, and 5 μl of 10× terminal deoxynucleotide transfer (TdT) buffer. Incubate the reaction for 1 h at 42 °C and then at 70 °C for 10 min.
3. Denature the reaction at 95 °C for 5 min and then rapidly cool the reaction on ice.
4. Add 0.4 μl of biotinylated ddATP and then 2 μl of TdT buffer. Incubate the reaction for 1 h at 37 °C and then at 70 °C for 10 min.
5. Digest the reaction with 1 μl of USER enzyme and incubate at 37 °C for 30 min.
6. Use DEPC water to bring the volume of the reaction to 60 μl and then add 72 μl of AMPure beads. Incubate the reaction for 30 min at room temperature with intermittent agitation.
7. Collect the beads using a magnetic stand and wash them twice with 500 μl of 70 % ethanol. Air-dry the beads for 5–10 min.
8. Resuspend the beads in 20 μl of TE buffer. Place the samples on the magnetic stand for 5 min and then remove the supernatant (*see* **Note 3**).

3.6 Sequencing

1. Sequence samples with the Helicos Single Molecule Sequencer (*see* **Note 4**).

3.7 Validation

1. Perform PCR reactions with Phusion High-Fidelity PCR Kit following the manufacturer's protocol.
2. Clean 7.5 μl of PCR sample with 3 μl of ExoSAP-IT following the manufacturer's protocol.
3. Sanger sequence the samples (*see* **Note 5**).

3.8 Bioinformatics and Machine Learning Methods

1. Align standard sequences.

   Align SMS reads from WT and ADAR deficient flies to DM3 reference genome. Perform alignments with the indexDP genomic aligner. Realign RNAseq reads that do not match the genome by setting the score for A-to-G or T-to-C substitutions (reference → read) to be the same as for the matching bases. Combine these alignments with the standard alignments for downstream processing. Capture all possible alignments for each RNAseq read and filter them with a minimum normalized score. Remove all reads mapping to rDNA or chrM (*see* **Note 6**).
2. Use basic filters for editing site discovery (Table 2) (*see* **Note 6**).

Table 2

Basic filters for editing site discovery

First-round filters	Second-round filters
Minimal number of reads with A-to-G substitutions ≥ 2	Minimal number of reads with A-to-G substitutions ≥ 4
Proportion of G substitutions at the candidate site G/(G + A + T + C + gap) ≥ 0.01	Average length of reads containing A-to-G substitutions at the position ≥ 31 bases
Removal of single-nucleotide polymorphisms (SNPs)	Removal of reads mapping to unmapped heterochromatic scaffolds (chrUextra)
Removal of A-to-G substitutions due to mis-mapping of reads spanning exon-exon junctions	Removal of sites with a possibility of adjacent single-nucleotide polymorphisms (SNPs)
Removal of candidate RNA-editing sites with two or more A-to-G substitutions in reads from ADAR-deficient flies	Removal of antisense sites except if they corresponded to an annotated antisense transcript

3. Use machine learning algorithms.

   Train the machine learning models on true positive (TP) and true negative (TN) sites obtained from Sanger sequencing validation. Assess the quality of the models using the following criteria: specificity (TN/N), sensitivity (TP/P), AUC (the area under receiver operating characteristic curve), and positive predictive value (TP/(TP + FP)). Furthermore, assess the performance of generated models through the ROCR package. Generate subsequent predictive models using the R environment for statistical computing. Finally, partition the data through extensive training and tuning of the models using the classification and regression training (caret) package for R (*see* **Note 7**).

4 Notes

1. For ribosomal depletion the manufacturer's protocol was modified as follows: Primers complementary to the *Drosophila* 18S and 28S rRNA transcript were designed to have a 5′-biotin. First, the primers were resuspended at 1000 μM and an equimolar mastermix prepared. A total of 2000 pmol oligo was added to 19 μl DEPC water and the mixture incubated at 95 °C for 5 min and then rapidly cooled on ice. A total of 8 μl of 10 mM ddNTPs, 4 μl 2.5 mM $CoCl_2$, 10× TdT buffer, and 3 μl Terminal Transferase were added to the primers and incubated at 37 °C for 1 h, followed by an additional incubation at 70 °C for 10 min. The primers were then cleaned twice on Performa

DTR cartridges following the manufacturer's protocols. 2.5 μl of the prepared primer mixture was added to the total RNA samples prior to hybridization, when the RiboMinus probe was added. The exact sequence of the *Drosophila* 18S and 28S rRNA complementary primers can be found in St. Laurent et al. [33].

2. Upon completion of cDNA synthesis RNA was eliminated by adding 1 μl RNAseH. We modified the manufacturer's protocol to include the addition of 1 μl RNAseIf as well and incubated the mixture at 37 °C for 30 min. Furthermore, the resulting cDNA was then purified by the serial use of two Performa Gel Filtration Columns and quantified.
3. An additional 20 μl elution was performed and pooled with the first sample.
4. 20 μl of samples were hybridized to the HeliScope flow cell at a loading concentration of 100–350 pM.
5. PCR and Sanger sequencing primers were designed with BatchPrimer 3 (BatchPrimer 3: a high-throughput Web application for PCR and sequencing primer design).
6. More detailed descriptions of computational methods implemented for sequencing alignments as well as information on the basic filters used for editing site discovery can be found in St. Laurent et al. [33].
7. Data used for the testing and training of machine learning models, description of variables, and additional details of the machine learning algorithms implemented for editing site discovery can be found in St. Laurent et al. [33].

Acknowledgements

The authors wish to thank members of the Reenan lab for helpful discussions and suggestions, especially Michael Fuchs, Natalie Palaychuk, and Ali Rezaei. The authors are grateful to Cindi Staber, Alexander Leydon, Mustafa Talay, and Alissa Trepman for useful comments.

References

1. Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem 71:817–846
2. Nishikura K (2010) Functions and regulation of RNA editing by adar deaminases. Annu Rev Biochem 79:321–349
3. Savva Y, Rieder LE, Reenan R (2012) The adar protein family. Genome Biol 13:252
4. Basilio C, Wahba AJ, Lengyel P, Speyer JF, Ochoa S (1962) Synthetic polynucleotides and the amino acid code, v. Proc Natl Acad Sci U S A 48:613–616
5. Hoopengardner B, Bhalla T, Staber C, Reenan R (2003) Nervous system targets of RNA editing identified by comparative genomics. Science 301:832–836

6. Bhalla T, Rosenthal JJC, Holmgren M, Reenan R (2004) Control of human potassium channel inactivation by editing of a small mRNA hairpin. Nat Struct Mol Biol 11:950–956
7. Rosenthal JJC, Seeburg PH (2012) A-to-I RNA editing: effects on proteins key to neural excitability. Neuron 74:432–439
8. Lehmann KA, Bass BL (2000) Double-stranded RNA adenosine deaminases adar1 and adar2 have overlapping specificities. Biochemistry 39:12875–12884
9. Savva YA, Jepson JEC, Chang Y-J, Whitaker R, Jones BC et al (2013) RNA editing regulates transposon-mediated heterochromatic gene silencing. Nat Commun 4:2745
10. Kumar M, Carmichael GG (1997) Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc Natl Acad Sci U S A 94:3542–3547
11. Bass BL, Weintraub H, Cattaneo R, Billeter MA (1989) Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA. Cell 56:331
12. Tonkin LA, Saccomanno L, Morse DP, Brodigan T, Krause M, Bass BL (2002) RNA editing by adars is important for normal behavior in *Caenorhabditis elegans*. EMBO J 21:6025–6035
13. Palladino MJ, Keegan LP, O'Connell MA, Reenan RA (2000) A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. Cell 102:437–449
14. Wang Q, Miyakoda M, Yang W, Khillan J, Stachura DL et al (2004) Stress-induced apoptosis associated with null mutation of adar1 RNA editing deaminase gene. J Biol Chem 279:4952–4961
15. Higuchi M, Maas S, Single FN, Hartner J, Rozov A et al (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme adar2. Nature 406:78–81
16. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T et al (2012) Landscape of transcription in human cells. Nature 489:101–108
17. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC et al (2005) The transcriptional landscape of the mammalian genome. Science 309:1559–1563
18. Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet 15:7–21
19. Serganov A, Serganov A, Patel DJ, Patel DJ (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet 8:776–790
20. Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y et al (2012) Global analysis of RNA secondary structure in two metazoans. Cell Rep 1:69–82
21. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY (2011) Understanding the transcriptome through RNA structure. Nat Rev Genet 12:641–655
22. Lukong KE, Chang K, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends Genet 24:416–425
23. Slotkin W, Nishikura K (2013) Adenosine-to-inosine RNA editing and human disease. Genome Med 5:105
24. Rice GI, Kasher PR, Forte GMA, Mannion NM, Greenwood SM et al (2012) Mutations in adar1 cause *Aicardi-Goutières* syndrome associated with a type I interferon signature. Nat Genet 44:1243–1248
25. Li M, Wang IX, Li Y, Bruzel A, Richards AL et al (2011) Widespread RNA and DNA sequence differences in the human transcriptome. Science 333:53–58
26. Kleinman CL, Majewski J (2012) Comment on "widespread RNA and DNA sequence differences in the human transcriptome". Science 335:1302
27. Lin W, Piskol R, Tan MH, Li JB (2012) Comment on "widespread RNA and DNA sequence differences in the human transcriptome". Science 335:1302
28. Pickrell JK, Gilad Y, Pritchard JK (2012) Response to comment on "widespread RNA and DNA sequence differences in the human transcriptome". Science 335:1302
29. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM et al (2011) The developmental transcriptome of *Drosophila melanogaster*. Nature 471:473–479
30. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P et al (2013) Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 10:128–132
31. Rodriguez J, Menet JS, Rosbash M (2012) Nascent-seq indicates widespread cotranscriptional RNA editing in *Drosophila*. Mol Cell 47:27–37
32. Sugden LA, Tackett MR, Savva YA, Thompson WA, Lawrence CE (2013) Assessing the validity and reproducibility of genome-scale predictions. Bioinformatics 29:2844–2851
33. St Laurent G, Tackett MR, Nechkin S, Shtokalo D, Antonets D et al (2013) Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in *Drosophila*. Nat Struct Mol Biol 20:1333–1339

34. Qu W, Shen Z, Zhao D, Yang Y, Zhang C (2009) Mfeprimer: multiple factor evaluation of the specificity of PCR primers. Bioinformatics 25:276–278
35. Ramaswami G, Lin W, Piskol R, Tan MH, Davis C, Li JB (2012) Accurate identification of human *Alu* and non-*Alu* RNA editing sites. Nat Methods 9:579–581
36. Peng Z, Cheng Y, Tan BC-M, Kang L, Tian Z et al (2012) Comprehensive analysis of RNA-seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol 30:253–260
37. Sam LT, Lipson D, Raz T, Cao X, Thompson J et al (2011) A comparison of single molecule and amplification based sequencing of cancer transcriptomes. PLoS One 6:e17305
38. Raz T, Kapranov P, Lipson D, Letovsky S, Milos PM, Thompson JF (2011) Protocol dependence of sequencing-based gene expression measurements. PLoS One 6:e19287
39. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L et al (2007) Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. Nature 450:219–232

Chapter 16

Nucleotide-Level Profiling of m^{5}C RNA Methylation

Tennille Sibbritt, Andrew Shafik, Susan J. Clark, and Thomas Preiss

Abstract

Mapping the position and quantifying the level of 5-methylcytosine (m^{5}C) as a modification in different types of cellular RNA is an important objective in the emerging field of epitranscriptomics. Bisulfite conversion has long been the gold standard for detection of m^{5}C in DNA but it can also be applied to RNA. Here, we detail methods for bisulfite treatment of RNA, locus-specific PCR amplification and detection of candidate sites by sequencing on the Illumina MiSeq platform.

Key words 5-Methylcytosine, Epitranscriptomics, Bisulfite conversion, Next-generation sequencing, MiSeq

1 Introduction

Cellular RNAs can be richly modified with more than one hundred known, chemically and structurally distinct nucleoside modifications [1–3]. The emerging field of epitranscriptomics [4–6] is enabled by the development of high-throughput mapping methods for RNA modifications, typically based on a next-generation sequencing (NGS) readout. Transcriptome-wide positions of 5-methylcytosine (m^{5}C) [7], N6-methyladenosine [8–10], and pseudouridine [11–13] have each been reported in this way. To detect m^{5}C in RNA, a range of methods have been developed, including the direct (meRIP [14]) or indirect (aza-IP [15], miCLIP [16]) immunoprecipitation of methylated RNA. Of particular interest here, the bisulfite conversion approach in popular use for DNA methylation detection has also been adapted to RNA [7, 17–19]. Bisulfite conversion of nucleic acids takes advantage of the differential chemical reactivity of m^{5}C compared to unmethylated cytosines; unmethylated cytosines are deaminated to uracil while m^{5}C remains as a cytosine.

We recently adapted an RNA bisulfite conversion method [17] for a NGS-based transcriptome-wide readout and mapped thousands of novel candidate m^{5}C sites in a variety of RNA biotypes,

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_16,

including mRNA [7]. Here, we detail our protocols for RNA bisulfite conversion and locus-specific, semiquantitative PCR-based detection of non-converted sites.

Sequencing of PCR amplicons is conveniently done on the Illumina MiSeq platform, as this affords multiplexing of multiple distinct amplicons while still achieving ample read depth for estimating the proportion of m^5C at targeted positions. For instance, each of the 24 indexed adaptors from the TruSeq DNA LT Sample Prep Kit could be assigned to a separate cellular RNA source material, and multiple RNA loci/PCR amplicons per sample could be included in the sequencing library, potentially generating hundreds of independent quantitative measurements of the m^5C level in a single MiSeq run (Fig. 1).

2 Materials

Prepare all solutions using DNase- and RNase-free H_2O and analytical grade reagents. Prepare and store all reagents at room temperature unless indicated otherwise. Diligently follow all safety and waste disposal regulations when performing experiments.

Prepare and perform bisulfite conversion, cDNA synthesis, and PCR amplification experiments in a PCR and plasmid-free area.

2.1 In Vitro Transcription Components

1. pRL Renilla Luciferase Reporter Vectors (pRL-TK) (Promega).
2. MEGAScript® T7 Kit (Life Technologies).
3. TURBO™ DNase (Life Technologies).
4. Phase Lock Gel Heavy (1.5 mL) (5 Prime).
5. UltraPure™ Phenol:Water (3.75:1 v/v) (Life Technologies).
6. Chloroform.
7. Glycogen (5 mg/mL) (Life Technologies).
8. Agilent RNA 6000 Nano Kit (Agilent).

2.2 Sodium Bisulfite Conversion Components

1. *Sodium bisulfite solution*: 40 % (w/v) sodium metabisulfite, 0.6 mM hydroquinone, pH 5.1.

   0.6 M Hydroquinone: Weigh 66 mg hydroquinone and place into a 1.5 mL tube. Add H_2O to 1 mL and cover in foil to protect from light. Place in an orbital shaker to dissolve (*see* **Note 1**).

   40% (w/v) sodium bisulfite: Dissolve 8 g sodium metabisulfite in 20 mL H_2O in a 50 mL falcon tube and vortex until it completely dissolves.

   Add 20 µL 0.6 M hydroquinone to the 40 % sodium bisulfite solution, vortex, and adjust pH to 5.1 with 10 M NaOH (*see* **Note 2**). Filter the solution through a 0.2 µm filter. Cover in foil to protect from light.

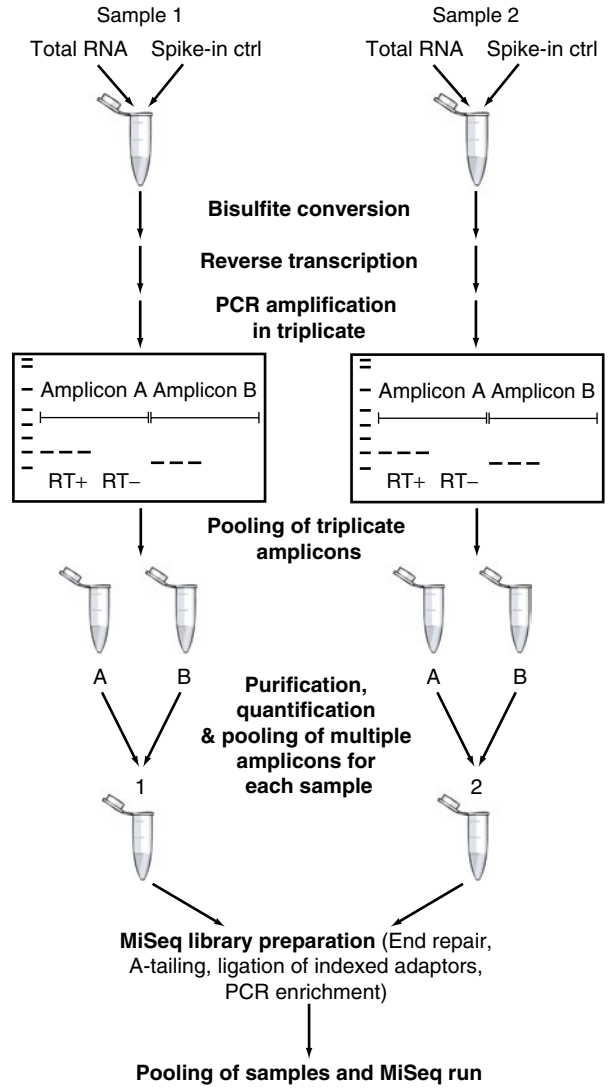

Fig. 1 Protocol overview, showing workflow and pooling strategy for effective sequencing. Total RNA spiked with the R-Luc in vitro transcript is bisulfite converted. The bisulfite-converted RNA is reverse transcribed and candidates of interest, as well as the positive (tRNA$^{Asp(GUC)}$ and tRNA$^{Leu(CAA)}$) and negative (Renilla Luciferase in vitro transcript) controls, are PCR amplified in triplicate for each sample to minimize PCR amplification bias. The triplicate amplicons are then pooled for each sample and subjected to purification. The purified amplicons are quantified and an equal amount of each amplicon is pooled for each sample. Following this, library preparation is performed using the TruSeq DNA LT Sample Prep Kit, which involves end repair and A-tailing of the amplicons, ligation of the indexed adaptors to the amplicons to enable multiplexing of samples, and enrichment of the libraries by PCR. Each library is then pooled, a PhiX control library is spiked in, and the libraries are subjected to sequencing on the MiSeq platform

2. 1 M Tris–HCl, pH 9.0.
3. Micro Bio-Spin® P-6 Gel Columns, Tris buffer (Bio-Rad).
4. Mineral oil.
5. 100 % ethanol.
6. 75 % ethanol.
7. 3 M sodium acetate, pH 5.2.
8. 5 mg/mL glycogen (Life Technologies).

2.3 cDNA Synthesis Components

1. SuperScript III® Reverse Transcriptase Kit (Life Technologies).
2. 10 mM mixed dNTPs.
3. 20× random primer mix: 35 μM hexamers, 25 μM T12VN.
   Oligo sequence for hexamers: NNNNNN
   Oligonucleotide sequence for T12VN: NVTTTTTTTTTTTT

2.4 PCR Amplification Components

1. Platinum® Taq DNA Polymerase Kit (Life Technologies).
2. 10 mM mixed dNTPs.

2.5 Agarose Gel Electrophoresis and PCR Purification Components

1. Seakem® LE Agarose (Lonza).
2. EZ-vision® Three DNA Dye and Buffer 6× (Amresco).
3. 1 Kb Plus DNA ladder (Life Technologies).
4. 1× TAE buffer: Make up 50× TAE buffer by combining 424 g Tris base, 57.1 mL acetic acid, and 100 mL 0.5 M EDTA (pH 8.0) and make up to 1 L in H_2O. To make 1× TAE buffer, combine 40 mL 50× TAE buffer with 1.96 L H_2O.
5. Wizard SV Gel and PCR Clean-Up System (Promega).

2.6 Library Preparation Components

1. TruSeq DNA LT Sample Prep Kit v1/v2 (Illumina).
2. MinElute PCR Purification Kit (Qiagen).
3. Agencourt AMPure XP beads (Beckman Coulter).
4. Tween 20 (Sigma).
5. EBT buffer: 10 mM Tris–HCl (pH 8.5) and 0.1 % Tween 20. Add 19.8 mL H_2O to 0.2 mL 1 M Tris–HCl (pH 8.5), and then add 20 μL Tween 20 to the solution (*see* **Note 3**). Vortex solution thoroughly to ensure that Tween 20 is mixed throughout the solution.
6. Library dilution buffer: 10 mM Tris–HCl (pH 8.0) and 0.05 % Tween 20. Add 19.8 mL H_2O to 0.2 mL 1 M Tris–HCl (pH 8.0) followed by 10 μL Tween 20. Vortex solution thoroughly to ensure that Tween 20 is mixed throughout the solution.
7. Library Quantification Kit-Illumina/Universal (Kapa Biosystems).
8. 0.2 M NaOH.
9. MiSeq Reagent Kit v2 (300 cycles) (Illumina) (*see* **Note 4**).

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 RNA Extraction and DNase Treatment

Total cellular RNA is extracted directly from adherent cells with 1 mL of Tri Reagent® as per the manufacturer's protocol. RNA is then treated with TURBO™ DNase as per the manufacturer's protocol.

3.2 Generation of the Renilla Luciferase (R-Luc) In Vitro Transcript Spike-In Control

1. Linearise the pRL-TK vector using *Bam*HI according to the MEGAScript® T7 Kit protocol.
2. Perform in vitro transcription according to the MEGAScript® T7 Kit protocol using 1 μg linearized DNA. An incubation period of 4 h at 37 °C with the kit components is sufficient.
3. Add 2 U TURBO™ DNase and incubate at 37 °C for 30 min.
4. Transfer the reaction to a Phase Lock Gel Heavy (1.5 mL) tube and make the volume of the reaction up to 100 μL with H_2O.
5. Add an equal volume of UltraPure™ Phenol:Water (3.75:1 v/v) and chloroform, shake vigorously for 15 s, and centrifuge at 16,000 × *g* for 5 min.
6. Add the same volume of chloroform as **step 5** to the tube, shake vigorously for 15 s, and centrifuge at 16,000 × *g* for 5 min again.
7. Transfer the aqueous phase to a clean 1.5 mL tube. Add 1/10 volume 3 M sodium acetate, 3 volumes of 100 % ethanol, and 1 μL glycogen (5 mg/mL), vortex, and precipitate the RNA overnight at −80 °C.
8. Centrifuge RNA at 17,000 × *g* at 4 °C for 30 min and carefully remove the supernatant.
9. Add 1 mL 75 % ethanol to the RNA, invert ~5 times and centrifuge at 7500 × *g* at 4 °C for 5 min (*see* **Note 5**).
10. Carefully remove the supernatant and let the pellet air-dry for 10–15 min (*see* **Note 6**).
11. Resuspend the RNA in H_2O.
12. Treat 10 μg in vitro transcript with 2 U TURBO™ DNase according to the manufacturer's protocol at 37 °C for 30 min to remove any residual template DNA.
13. Assess the size and integrity of the in vitro transcript using a RNA 6000 Nano Chip on the Agilent® 2100 Bioanalyzer according to the manufacturer's protocol.

3.3 Bisulfite Conversion of RNA

1. Add 1/1000 R-Luc in vitro transcript to 2–4 μg DNase-treated RNA (*see* **Note 7**). The combined volume of the RNA sample and in vitro transcript should be <20 μL.
2. Denature RNA by heating to 75 °C for 5 min in a heat block.
3. Preheat sodium bisulfite solution to 75 °C, add 100 μL to the RNA, vortex thoroughly, and briefly spin in a microcentrifuge.
4. Overlay the reaction mixture with 100 μL mineral oil (*see* **Note 8**). Cover the tube in aluminium foil to protect the reaction mixture from light.
5. Incubate at 75 °C for 4 h in a heat block.
6. About 10 min before the bisulfite conversion reaction is complete, prepare two Micro Bio-Spin® P-6 Gel Columns in Tris buffer per reaction by allowing to drain into a collection tube by gravity (*see* **Note 9**). Discard the flow-through, place the column back into the collection tube, and centrifuge at 1000 × *g* for 2 min. Transfer each column to a clean 1.5 mL tube.
7. Remove the bisulfite reaction mixture from the heat and gently transfer the bottom layer containing the sodium bisulfite/RNA mixture to the Micro Bio-Spin® P-6 Gel Column in Tris buffer (*see* **Note 10**).
8. Centrifuge at 1000 × *g* for 4 min.
9. Carefully transfer the eluate into the second Micro Bio-Spin® P-6 Gel Column placed in the 1.5 mL tube and repeat **step 8**.
10. Discard the column, add an equal volume of 1 M Tris–HCl (pH 9.0) to the second eluate, vortex, spin briefly, and then overlay with 150 μL mineral oil. Cover the tube in aluminium foil to protect the reaction mixture from light.
11. Incubate at 75 °C for 1 h in a heat block.
12. Transfer the bottom layer containing the RNA to a clean 1.5 mL tube. Precipitate overnight and resuspend bisulfite-converted RNA in H_2O as described in **steps 7–11** in Subheading 3.2 (*see* **Notes 11** and **12**).

3.4 cDNA Synthesis

1. Incubate 200–300 ng bisulfite-converted RNA, 1 μL 20× random primer mix, 1 μL of 1 mM dNTP mix, and H_2O to a final volume of 13 μL at 65 °C for 5 min to denature the RNA.
2. Reverse transcribe the bisulfite-converted RNA using the standard manufacturer's protocol. It is essential to include RT– controls for each sample as the primers are not necessarily designed to span exon-exon junctions. Use 1 μL H_2O instead of SuperScript III® reverse transcriptase for the RT– control.
3. Once the reaction is complete, dilute cDNA 1:10 in H_2O for bisulfite PCR amplification.

3.5 Bisulfite PCR Primer Design

1. For the amplification of the R-Luc in vitro transcript, which does not contain m5C sites, design the primers such that they avoid areas of bisulfite-converted cytosines (*see* **Note 13**) (Fig. 2a).

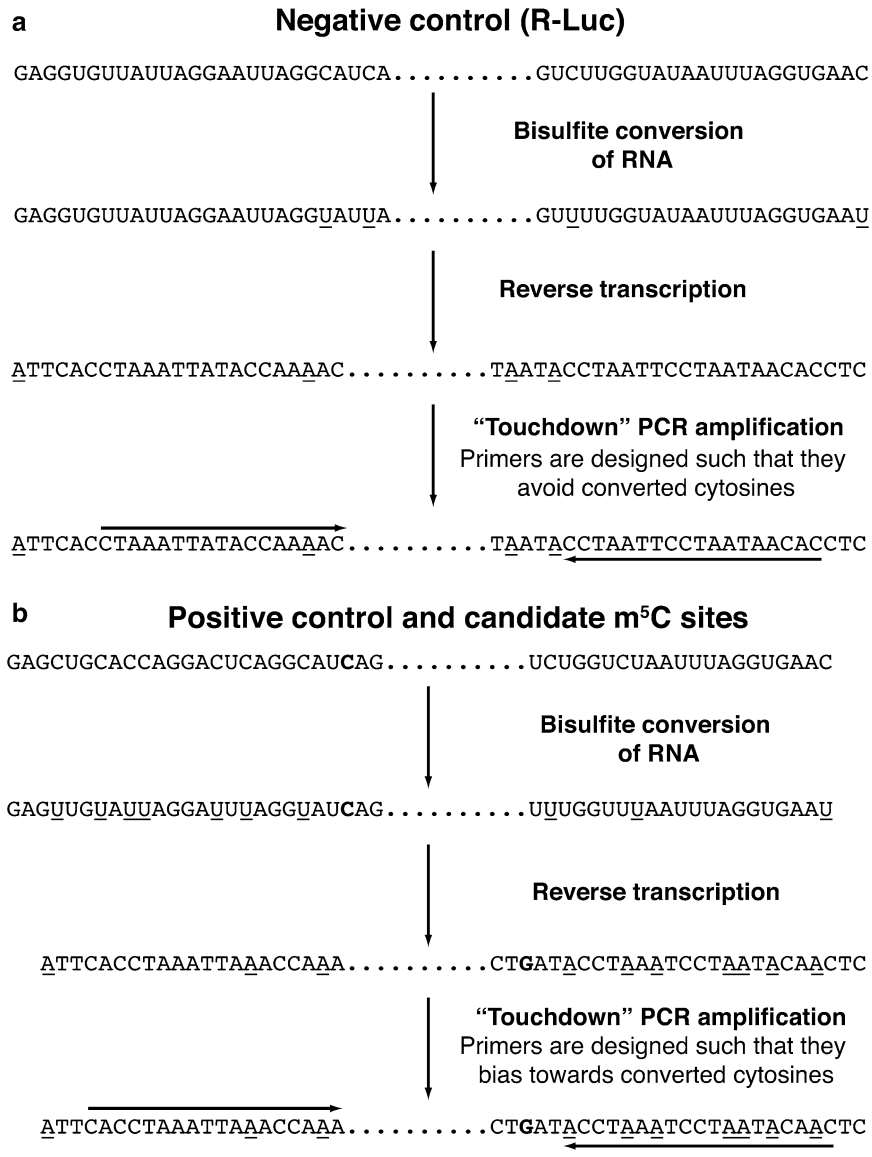

Fig. 2 Schematic demonstrating the bisulfite conversion of RNA, reverse transcription, primer design, and PCR amplification. (**a**) Primers designed for the R-Luc in vitro transcript avoid areas of converted cytosines to prevent preferential amplification of converted sequences, which may falsely indicate efficient bisulfite conversion. (**b**) Primers designed for the tRNA controls and validation of candidate m5C sites span areas containing converted cytosines to preferentially amplify converted sequences. *Underlined* bases represent converted (unmethylated) cytosines and *bold bases* represent m5C. Primers were designed to amplify products that were 70–200 bp

2. For the amplification of positive control transcripts (*see* **Note 14**) and transcripts containing candidate m^5C sites, design the primers such that they span regions containing converted cytosines to avoid preferentially amplifying non-converted sequences (Fig. 2b).
3. Smaller amplicon sizes are desired to reduce amplification of non-converted cytosines due to strong secondary structure (*see* **Note 15**). Amplicons can be design such that they are 70–200 bp in length.
4. To ensure the designed primers uniquely amplify the region of interest, check primer specificity using the BiSearch webserver [20].

3.6 PCR Amplification and Pooling of Amplicons

1. For a 25 μL PCR reaction, add 0.1 μL Platinum® Taq DNA Polymerase to 1 μL diluted cDNA, 2.5 μL 10× buffer (without $MgCl_2$), 0.5 μL 10 mM dNTPs, 0.75 μL 50 mM $MgCl_2$, 0.25 μL 10 μM forward primer, 0.25 μL 10 μM reverse primer, and 19.65 μL of H_2O. Perform PCR for each cDNA sample in triplicate to reduce the potential for PCR amplification bias.
2. Mix gently, spin briefly, and place into a thermal cycler.
3. Perform a "touchdown" PCR program, which is executed in two phases (Table 1) (*see* **Note 16**).
4. Perform standard 2 % agarose gel electrophoresis of all amplicons along with 5 μL 1 Kb Plus DNA ladder. Run the gel at 100 V for ~40 min in 1× TAE.

Table 1

"Touchdown" PCR cycling conditions for the amplification of candidates by Platinum® Taq DNA polymerase

Stage		Temperature	Time
Initial denaturation		94 °C	2 min
Phase I			
Denaturation		94 °C	30 s
Annealing		Cycle from $Tm_{highest}$ + 5 °C to Tm_{lowest} – 5 °C	30 s
Extension		72 °C	15–30 s (*see* **Note 17**)
Phase II			
25–45 cycles	Denaturation	94 °C	30 s
	Annealing	Tm_{lowest} – 5 °C	30 s
	Extension	72 °C	15–30 s
Final extension		72 °C	5 min
Hold		4 °C	Forever

5. Pool the triplicate amplicons (*see* **Note 18**) and purify them using the Wizard SV Gel and PCR Clean-Up System according to the manufacturer's protocol (*see* **Note 19**).
6. Quantify each set of pooled triplicate amplicons using the Qubit dsDNA BR Assay Kit according to the manufacturer's protocol (*see* **Note 20**).
7. The MiSeq protocol enables the sequencing of multiple candidates on a single sequencing run. Pool 20–30 ng of each amplicon per sample into a single 1.5 mL tube.
8. If the volume is >55 μL, concentrate it down to ~55 μL or less using a vacuum concentrator and make the volume up to 55 μL in H_2O. If the volume is <55 μL, make it up to 55 μL in H_2O (*see* **Note 21**).

3.7 MiSeq Amplicon Sequencing Library Preparation

This protocol has been adapted from the TruSeq DNA LT Sample Preparation Guide.

1. Add 10 μL End Repair Control and 40 μL End Repair Mix to 50 μL of each pooled set of amplicons (*see* **Note 22**), gently pipette the entire volume up and down ten times, and incubate at 30 °C for 30 min in a thermal cycler (*see* **Note 23**).
2. For libraries containing amplicon sizes <100 nt, purify them using the MinElute PCR Purification Kit, as per the manufacturer's instructions (*see* **Note 24**). For libraries containing amplicon sizes >100 nt, purify the library using AMPure XP beads as detailed below (*see* **Note 25**).
3. Vortex AMPure XP beads to ensure even distribution. Dilute 136 μL beads in 24 μL H_2O, add to the End Repair reaction mixture, gently pipette the entire volume up and down ten times to mix, and incubate at room temperature for 15 min.
4. Place the tubes on a magnetic rack at room temperature for 5 min and then remove the supernatant.
5. Keep the tubes in the magnetic rack and wash the beads twice for 30 s by gently adding 200 μL of freshly made 80 % ethanol to the tubes without disturbing the beads. Allow to air-dry at room temperature for 15 min.
6. Remove the tubes from the magnetic rack and resuspend the beads in 17.5 μL resuspension buffer. Incubate at room temperature for 2 min.
7. Place the tubes on the magnetic rack for 5 min then transfer 15 μL of the cleared supernatant to a new 0.5 mL tube.
8. Add 2.5 μL A-tailing Control and 12.5 μL A-tailing Mix to 15 μL supernatant, pipette the entire volume of the reaction mixture up and down ten times, and incubate at 37 °C for 30 min in a thermal cycler.

9. Remove the tubes from the thermal cycler and add 2.5 μL ligation control, 2.5 μL ligation mix and 2.5 μL DNA adaptor index to the 30 μL reaction mixture (*see* **Note 26**).
10. Pipette the entire volume of the reaction mixture up and down ten times and centrifuge at 280 × *g* for 1 min.
11. Incubate at 30 °C for 15 min in a thermal cycler.
12. Remove the reaction mixture from the thermal cycler and add 5 μL stop ligation buffer. Pipette the entire volume up and down ten times to mix.
13. Purify the reaction mixture using 42.5 μL *undiluted* AMPure XP beads as described in **steps 3**–7, resuspending the beads in 52.5 μL resuspension buffer (*see* **Note 27**).
14. Transfer 50 μL of the reaction mixture to a new 0.5 mL tube.
15. Purify the reaction mixture again using 50 μL *undiluted* AMPure XP as described in **steps 3**–7, resuspending the beads in 22.5 μL resuspension buffer.
16. Transfer 20 μL of the cleared supernatant to a new 0.5 mL tube.
17. Add 25 μL PCR Master Mix and 5 μL PCR Primer Cocktail to the 20 μL reaction mixture and pipette the entire volume up and down ten times to mix.
18. Place the reaction mixture into a thermal cycler and perform the PCR program as outlined in Table 2 (*see* **Note 28**).
19. Purify the enriched library using 50 μL *undiluted* AMPure XP beads, as described in **steps 3**–7, resuspending the beads in 32.5 μL resuspension buffer.
20. Transfer 30 μL of the cleared supernatant to a new 0.5 mL tube. The library can now be stored at −20 °C for up to 7 days.

3.8 Validation and Quantification of the Libraries

1. Dilute 1 μL library in 5 μL H_2O and dilute 2 μL pooled amplicons prior to library preparation in 4 μL H_2O (*see* **Note 22**).
2. To validate that each library preparation was successful and to determine the average library size, perform standard 2 %

Table 2

PCR cycling conditions for the enrichment of the MiSeq amplicon sequencing library

Stage		Temperature	Time
Initial denaturation		98 °C	30 s
10 cycles	Denaturation	98 °C	10 s
	Annealing	60 °C	30 s
	Extension	72 °C	30 s
Final extension		72 °C	5 min
Hold		4 °C	Forever

agarose gel electrophoresis of all pooled amplicons and libraries along with 5 μL 1 Kb Plus DNA ladder. Run the gel at 100 V for ~40 min in 1× TAE.

3. Quantify the libraries using the Qubit dsDNA BR Assay Kit, following the manufacturer's protocol.
4. Determine the concentration of each library in nM using the average fragment size determined from **step 2** and the concentration in ng/μL determined from **step 3**. The following formula may be used: $\frac{[\text{DNA}](\text{ng}/\mu\text{L})}{\text{Average fragment size (bp)} \times 649} \times 10^6$
5. Dilute the libraries to 50 nM with EBT buffer.
6. Perform the final quantification of each library by qPCR using the Library Quantification Kit-Illumina/Universal according to the manufacturer's protocol with minor modifications. The manufacturer's protocol recommends performing twofold serial dilutions of the libraries from 1:1000 to 1:8000 in library dilution buffer; however this range results in most samples being outside the range of the standard curve. Perform twofold serial dilutions of the libraries from 1:1000 to 1:32,000 to ensure they are within the range of the standard curve.
7. To calculate the concentration of each library from the qPCR, average the concentration determined for each library based on the dilutions within the range of the standard curve.

3.9 Preparation of the Sample Sheet

1. Use the Illumina Experiment Manager based on the manufacturer's protocol to prepare the sample sheet (*see* **Note 29**).

3.10 Dilution of the Libraries and Loading of the Cartridge

1. Dilute each library to 10 nM in EBT buffer based on the concentrations determined by the qPCR (*see* **Note 30**). From this point, keep the libraries on ice.
2. Pool 10 μL of each 10 nM library, and further dilute to 2 nM by adding 8 μL EBT buffer to 2 μL of the 10 nM pooled libraries.
3. Dilute the PhiX control library to 2 nM by adding 8 μL EBT buffer to 2 μL of the 10 nM PhiX control library (*see* **Note 31**).
4. Denature the pooled libraries and PhiX control library separately by adding 10 μL 0.2 M NaOH to 10 μL of the 2 nM libraries (*see* **Note 32**).
5. Vortex to mix and centrifuge at 280 × *g* for 1 min. Incubate at room temperature for 5 min.
6. Dilute the denatured pooled libraries and PhiX control library separately to 20 pM by adding 980 μL pre-chilled HT1 to 20 μL denatured libraries.
7. Dilute the 20 pM pooled libraries and PhiX control library separately to 10 pM by adding 500 μL pre-chilled HT1 to 500 μL 20 pM libraries.

8. Combine 100 μL of the 10 pM PhiX control library with 900 μL of the 10 pM pooled libraries and vortex to mix (*see* **Notes 33** and **34**).
9. Load 600 μL of the final sample into the cartridge. Ensure that air bubbles are removed by gently tapping the cartridge.
10. Perform the sequencing run according to the manufacturer's protocol.

3.11 Alignment of the MiSeq Data

To trim the Illumina® adaptor sequences that were ligated to the ends of amplicons to facilitate sequencing of the 150 bp paired-end reads, use Trimmomatic [21] in palindromic mode (*see* **Note 35**). Sequencing reads can be aligned with Bismark [22], using Bowtie2 internally and implementing the parameters bismark --non_directional --bowtie2 (*see* **Note 36**). As the reference sequences for the alignment, use the segment of RNA interrogated by sequencing prior to bisulfite conversion. The number of C and T calls at all C positions can be extracted from the aligned sequencing reads in order to determine the proportion of m^5C at a given cytosine (Fig. 3).

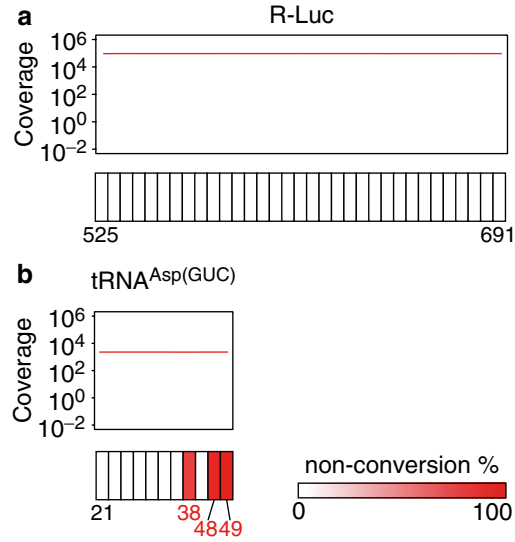

Fig. 3 Representative output of MiSeq amplicon sequencing showing a segment of the negative control R-Luc spike-in transcript and the positive control $tRNA^{Asp(GUC)}$. The *top panel* displays the coverage of the amplicon, which is relatively even across the amplicon. The heatmaps in the *bottom panel* display the cytosine non-conversion percentage. *Numbers* below the heatmaps represent the positions of cytosines relative to the start of the transcript. (**a**) A segment of the R-Luc in vitro transcript that does not exhibit non-conversion of cytosines, indicating the bisulfite conversion reaction was efficient. (**b**) A segment of endogenous HeLa cell $tRNA^{Asp(GUC)}$ displaying high levels of cytosine non-conversion (methylation) at the known m^5C sites (C38, C48, and C49)

4 Notes

1. Dissolving hydroquinone in H_2O can be time consuming, particularly in cooler temperatures. It is recommended that this is prepared first.
2. 10 M NaOH is added dropwise to the sodium bisulfite solution while mixing. Slightly less than 1 mL is required to adjust the pH to 5.1.
3. 1 M Tris–HCl (pH 8.5) can be obtained from Buffer EB from Qiagen purification kits. This can be used for the preparation of EBT buffer.
4. The MiSeq Reagent Kit v2 (300 cycles) provides 2×150 bp reads.
5. Do not vortex the RNA as this will increase the risk of RNA loss.
6. Air-drying the sample in a biohazard hood is best. Ensure that RNA does not completely dry as this will cause difficulties in the resuspension.
7. As the in vitro transcript will most likely be at high concentrations, it is best to perform a serial dilution in H_2O. Dilute to a concentration such that 3–4 μL of the in vitro transcript is added to the RNA samples for accurate pipetting.
8. It is best to tilt the 1.5 mL tube at a 45° angle and then slowly pipette the mineral oil directly on top of the reaction mixture.
9. Emptying of the P-6 gel column takes ~2 min. If the gel column does not empty by gravity, place the lid back onto the column and remove again.
10. Slowly and gently pipette the reaction mixture onto the gel bed. Avoid disturbing the gel bed. Minimize the transfer of mineral oil, although there will be traces which is unavoidable. The mineral oil will increase the A_{260}/A_{230} ratio to >2; however this does not hinder subsequent reactions.
11. ~500 ng is lost during this procedure, and we find that 13–15 μL of H_2O per 2 μg RNA used in the bisulfite conversion reaction results in concentrations of ~100 ng/μL.
12. For transcriptome-wide detection of m^{5}C (bsRNA-seq), confirmation of fragment size using the Agilent® 2100 Bioanalyzer RNA 6000 Nano Chip according to the manufacturer's protocol is required. This is not required for preparation of RNA for locus-specific sequencing using Sanger sequencing or the MiSeq platform.
13. Inefficient bisulfite conversion may result in unconverted cytosines, so it is necessary to ensure the primers are not biasing towards converted cytosines.

14. $tRNA^{Asp(GUC)}$ is known to contain m^5C sites at C38, C48, and C49, and $tRNA^{Leu(CAA)}$ is known to contain a m^5C site at C34. We have previously used these transcripts as positive controls for m^5C sites.
15. Longer amplicons increases the propensity of detecting non-converted cytosines in RNA exhibiting strong secondary structure. We have previously experienced this for amplicons derived from rRNA.
16. A "touchdown" PCR is performed to increase the specificity of the product. The first phase uses a higher annealing temperature, amplifying the specific product. The annealing temperature at each subsequent cycle is decreased by 1 °C to approximately 5 °C below the lowest primer melting temperature. In the second phase, a standard PCR protocol is implemented using the lowest annealing temperature used in the first phase.
17. Use an extension time of 30 s for amplicons of ~200 bp. For amplicons <100 bp, use an extension time of 15 s.
18. Sometimes all triplicates do not successfully amplify, and it may be necessary to optimise the PCR.
19. We recommend eluting the purified PCR products in 15–30 μL depending on the amount of amplified PCR product.
20. For quantification using the Qubit dsDNA BR Assay Kit, use 2 μL DNA.
21. After purification of the amplicons with the Wizard SV Gel and PCR Clean-Up System, residual ethanol may remain in the purified amplicons. We find that concentrating down the pooled amplicons even if there is <55 μL and addition of H_2O to 55 μL is best to remove as much ethanol as possible.
22. The remaining 5 μL of pooled amplicons is kept for the validation of the libraries.
23. For the library preparation, use 0.5 mL PCR tubes. However, if the thermal cycler or magnetic rack used cannot accommodate these sized tubes, it is possible to switch between 1.5 mL tubes and 0.2 mL tubes.
24. Agencourt AMPure XP beads purify DNA fragments >100 nt. Purification of the DNA fragments between 70 and 100 nt is achieved using the MinElute PCR Purification Kit. However, some loss of DNA fragments may occur using the column purification and there may be residual ethanol.
25. Remove Agencourt AMPure XP beads from the fridge for at least 30 min prior to use.
26. The indexed adaptors allow multiplexing of samples on a single sequencing run. Ensure that you read the TruSeq Sample Preparation Guide if you have low diversity libraries. Low-diversity libraries do not contain many different samples or amplicons. Some indexed adaptors are not compatible with

each other for low diversity libraries (i.e., <4 samples). This guide provides information on which indexed adaptors may be used for these libraries.

27. After the ligation of indexed adaptors, all DNA fragments will be >100 nt. As a result, Agencourt AMPure XP beads can be used.
28. Depending on the thermal cycler and ramp rate used, the PCR enrichment reaction can take 30–45 min.
29. The sample sheet is required to enter in the sample names and adaptor indices used for each sample. We have previously selected the "Other" as the category followed by "Fastq only" as the application for MiSeq amplicon sequencing. This generates fastq files only and also enables the deselection of adaptor trimming, allowing trimming and mapping to be performed separately. However, it is possible to select "small genome sequencing" as the category and "resequencing" as the application. This workflow includes adaptor trimming.
30. We have previously used 5 μL of each library for the dilution to 10 nM. Store the remaining libraries at –20 °C.
31. The PhiX library is added to the pooled libraries as a control for the sequencing run.
32. Prepare fresh 0.2 M NaOH for the denaturation of libraries.
33. Loading of 10 % PhiX control library is sufficient even for low-diversity libraries.
34. The concentration of the library to be loaded into the cartridge can vary. We have previously loaded between 7 and 9 pM. Underloading of the libraries can give cluster densities below the optimal range. Overloading of the libraries can give cluster densities above the optimal range, reducing the quality of the data. The optimal cluster density is 700–1000 K/mm^2.
35. Previously, bases at the 5′ end that were below the Phred quality score of 3 and bases at the 3′ end that were below the Phred quality score of 15 were removed. If any four base window within the read had a Phred quality score of <15, then the rest of the read was trimmed.
36. We have previously allowed a single mismatch in the alignment, and candidate m^{5}C sites in overlapped paired-end reads were only counted once.

Acknowledgements

We thank Wenjia Qu for helpful suggestions for the MiSeq library preparation protocol. We also thank Ulrike Schumann for helpful suggestions on this manuscript. This work was supported by an NHMRC grant (APP1061551) and a Senior Research Fellowship (514904) awarded to TP.

References

1. Cantara WA, Crain PF, Rozenski J et al (2011) The RNA modification database, RNAMDB: 2011 update. Nucleic Acids Res 39: D195–D201
2. Milanowska K, Mikolajczak K, Lukasik A et al (2013) RNApathwaysDB--a database of RNA maturation and decay pathways. Nucleic Acids Res 41:D268
3. Machnicka MA, Milanowska K, Osman OO et al (2012) MODOMICS: a database of RNA modification pathways--2012 update. Nucleic Acids Res 41:D262–D267
4. Saletore Y, Meyer K, Korlach J et al (2012) The birth of the Epitranscriptome: deciphering the function of RNA modifications. Genome Biol 13:175
5. Fu Y, He C (2012) Nucleic acid modifications with epigenetic significance. Curr Opin Chem Biol 16:516–524
6. Sibbritt T, Patel HR, Preiss T (2013) Mapping and significance of the mRNA methylome. WIREs RNA 4:397–422
7. Squires JE, Patel HR, Nousch M et al (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res 40:5023–5033
8. Meyer KD, Saletore Y, Zumbo P et al (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell 149:1635–1646
9. Dominissini D, Moshitch-Moshkovitz S, Schwartz S et al (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485:201
10. Schwartz S, Mumbach MR, Jovanovic M et al (2014) Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep 8:284–296
11. Schwartz S, Bernstein DA, Mumbach MR et al (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell 159: 148–162
12. Carlile TM, Rojas-Duran MF, Zinshteyn B et al (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515:143–146
13. Lovejoy AF, Riordan DP, Brown PO (2014) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in S. cerevisiae. PLoS One 9:e110799
14. Edelheit S, Schwartz S, Mumbach MR et al (2013) Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet 9:e1003602
15. Khoddami V, Cairns BR (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat Biotechnol 31:458–464
16. Hussain S, Sajini AA, Blanco S et al (2013) NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep 4:255–261
17. Gu W, Hurto RL, Hopper AK et al (2005) Depletion of Saccharomyces cerevisiae tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5) C. Mol Cell Biol 25:8191–8201
18. Schaefer M, Pollex T, Hanna K et al (2009) RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res 37:e12
19. Pollex T, Hanna K, Schaefer M (2010) Detection of cytosine methylation in RNA using bisulfite sequencing. Cold Spring Harb Protoc 2010:pdb.prot5505
20. Tusnady GE, Simon I, Varadi A et al (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. Nucleic Acids Res 33:e9
21. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
22. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27:1571–1572

Chapter 17

Probing *N*^{6}-methyladenosine (m^{6}A) RNA Modification in Total RNA with SCARLET

Nian Liu and Tao Pan

Abstract

Posttranscriptional *N*^{6}-methyladenosine (m^{6}A) RNA modification is indispensable for cell development and viability; however, functional investigation of m^{6}A biological function has been hindered by the lack of methods for its precise identification and quantitation. Here, we describe a method that accurately identifies m^{6}A position and modification fraction in human messenger RNA (mRNA) and long noncoding RNA (lncRNA) at single-nucleotide resolution, termed as "site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET)" (Fig. 1). This method combines two previously established techniques, site-specific cleavage and splint ligation, to probe the m^{6}A RNA modification status at any mRNA/lncRNA site in the total RNA pool.

Key words *N*^{6}-methyladenosine, SCARLET, RNA modification, Modification fraction, Single-nucleotide resolution, mRNA/lncRNA

1 Introduction

Discovered in the 1970s, m^{6}A is the most prevalent internal mRNA/lncRNA modification in eukaryotes, present on average in over three sites per mRNA molecule in mammals [1–6]. The m^{6}A/MeRIP-seq revealed the m^{6}A topology along mammalian transcript at ~100 nucleotides resolution [7–9]. In this chapter, we describe a protocol to determine the exact position as well as the modification fraction of mRNA/lncRNA modifications at single-nucleotide resolution without the need to isolate the target RNA. This method has been applied to reveal the m^{6}A status in human mRNA/lncRNAs, which provided information on the location and structural contexts of m^{6}A modification [10].

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_17,

2 Materials

Prepare all solutions using RNase-free water (prepared by autoclaving deionized water).

1. PerfectPure RNA cultured cell kit (5′).
2. GenElute mRNA miniprep kit (Sigma-Aldrich).
3. T4 PNK (USB).
4. Thermosensitive alkaline phosphatase, TAP (Thermo Scientific).
5. Crush and soak buffer, 50 mM Potassium Acetate, 200 mM KCl, pH 7.
6. RNase T1 (Thermo-Scientific).
7. RNase A (Sigma-Aldrich).
8. T4 DNA ligase (Thermo-Scientific).
9. Nuclease P1 (Sigma-Aldrich).
10. TLC cellulose plastic sheet, 20 × 20 cm (Merck).
11. TLC running buffer, isopropanol:HCl:water, 70:15:15, v/v/v.

3 Methods

The method is termed as "*s*ite-specific *c*leavage *a*nd *radioactive-labeling* followed by *l*igation-assisted *e*xtraction and *t*hin-layer chromatography (SCARLET)." SCARLET is composed of four main steps: site-specific cleavage at the target nucleotide site; radioactive labeling of the target nucleotide; splint-assisted ligation followed by RNase T1/A digestion; and thin-layer chromatography (TLC) (Fig. 1). We describe the SCARLET method in detail below in each of these four steps. Perform at room temperature (RT) unless specifically indicated.

When interested in the modification status (the presence of modification and the modification fraction) of the target nucleotide X along the target mRNA Y, we need to first design the chimeric oligos, splint oligos and ssDNA oligos according to the sequence of the target mRNA Y, as previously reported [10–14]. All chimeric oligos and DNA oligos can be ordered from IDT, and gel purified before use.

Before performing SCARLET, isolate total RNA from HeLa cells or other cell lines using PerfectPure RNA cultured cell kit (5′) according to the manual. Then, isolate polyadenylated RNA (polyA$^+$ RNA) from the total RNA sample via the GenElute mRNA miniprep kit (Sigma-Aldrich) according to the manual. Store the RNA at −80 °C until ready for SCARLET.

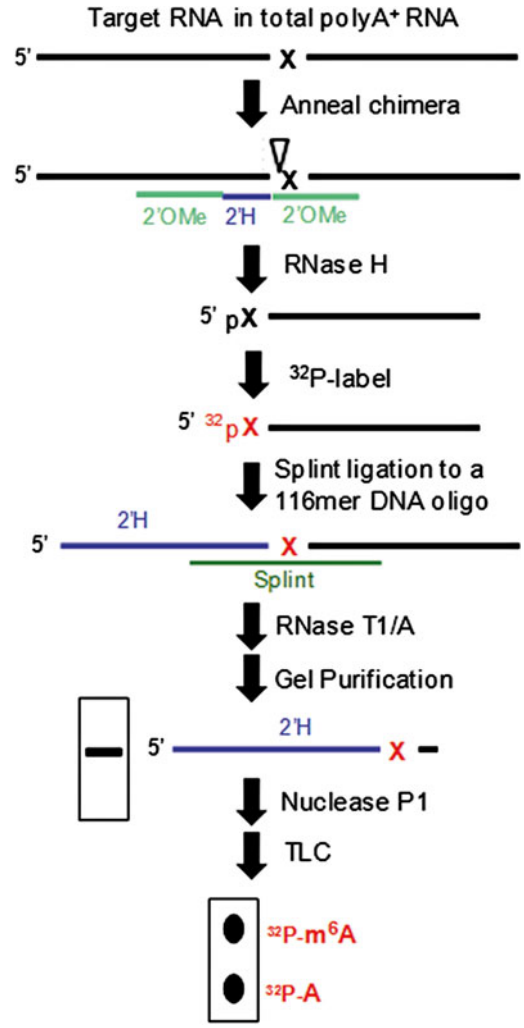

Fig. 1 Schematic diagram of SCARLET. SCARLET consists of four steps: site-specific cleavage at the target nucleotide site; radioactive labeling of the target nucleotide; splint-assisted ligation followed by RNase T1/A digestion; and thin-layer chromatography (TLC)

3.1 Site-Specific Cleavage at the Target Nucleotide Site

Day 1

1. Mix 1 μg polyA+ RNA with (3 pmol) corresponding chimeric oligo in a total volume of 3 μl 30 mM Tris–HCl, pH 7.5. Anneal the oligo to RNA by heating at 95 °C for 1 min, followed by incubation at room temperature (RT) for 3 min before putting on ice for the next step.
2. Add 1 μl 5× RNase H reaction mixture [2× T4 polynucleotide kinase buffer (T4 PNK, USB), 1 U/μl RNase H (Epicentre)] and 1 μl thermosensitive alkaline phosphatase (1 U/μl, TAP, Thermo Scientific) to the annealed RNA sample. Incubate at

44 °C for 1 h for site-specific cleavage and dephosphorylation at the 5′ end of nucleic acids.

3. Terminate the reaction by heating the reaction mixture at 75 °C for 5 min followed by immediate incubation on ice to inactivate RNase H and TAP.

3.2 Radioactive Labeling of the Target Nucleotide

1. Add 1 μl 6× T4 PNK reaction mixture [1× T4 PNK buffer, 6 U/μl T4 PNK (USB), 28 μCi/μl [γ-^{32}P]-ATP] to the mixture from above.
2. Incubate at 37 °C for 1 h.
3. Terminate the reaction by heating the mixture at 75 °C for 5 min followed by immediate incubation on ice to inactive the T4 PNK.

3.3 Splint-Assisted Ligation Followed by RNase T1/A Digestion

1. Add to the reaction mixture from above 1.5 μl the splint/ssDNA-116 oligo mixture (4 pmol splint oligos and 5 pmol ssDNA-116 oligos), and mix well.
2. Anneal the RNA samples, splint oligos, and ssDNA-116 oligos by heating the mixture at 75 °C for 3 min, followed by incubation at room temperature (RT) for 3 min before putting on ice for the next step.
3. Add 2.5 μl 4× ligation mixture [1.4× T4 PNK buffer, 0.27 mM ATP, 57 % DMSO, 1.9 U/μl T4 DNA ligase].
4. Incubate at 37 °C for 3.5 h for the splint ligation.
5. Terminate the reaction by mixing the reaction samples with equal volume of 2× RNA loading buffer (9 M urea, 100 mM EDTA, XC, and BPB dyes).
6. Add 1 μl RNase T1/A mixture (160 U/μl RNase T1, 0.16 mg/ml RNase A in distilled water), and mix well.
7. Incubate at 37 °C overnight (~16 h) to ensure complete RNase digestion.

Day 2

8. Spin down the reaction mixture. Load all samples to a pre-run 10 % urea denaturing PAGE gels (0.8 mm double-thick gel). Run the bromophenol dye to the bottom.
9. Disassemble the gel electrophoresis equipment, wrap the gel with plastic film, and expose the gel to a blanked phosphorimager screen. To get clear phosphorimaging figures with visible target bands, the exposure time varies from 10 to 30 min, depending on the radioactive signal strength (Fig. 2).
10. Visualize and print the phosphorimager figure in the actual size. Put the printed figures under the gel for localization of the target bands.

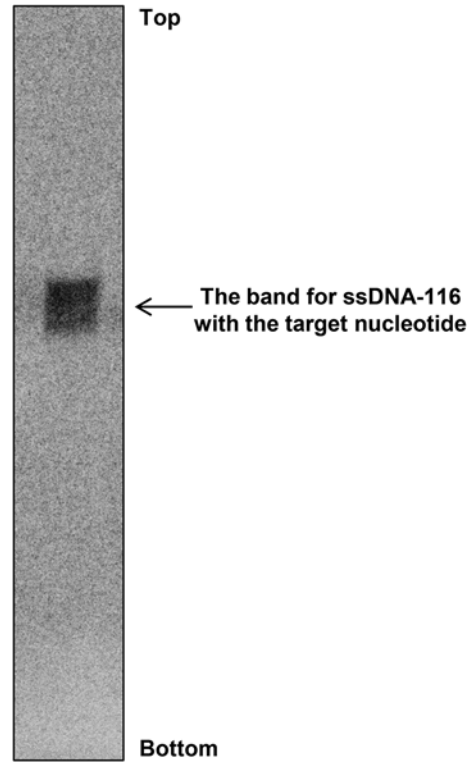

Fig. 2 Denaturing PAGE showing the band for ssDNA-116 oligos with ligated radioactive target nucleotide. This band was derived from the SCARLET experiment with the 2577-A site on the MALAT1 lncRNA (NR_002819) from HFF-1 cells

11. Cut the bands with flame-sterilized blades from the gel and transfer the cut gel slices into a clean 1.5 ml low-adhesion, plastic tube.
12. Add 0.4 ml crush and soak buffer to the tube, and invert and rotate the tube for 4 h at RT.
13. Transfer the crush and soak buffer containing the target oligos out to new low-adhesion tubes, and add 2.7× vol. of pure ethanol. Mix well, freeze at −20 °C for at least 1 h, and then precipitate the nucleic acid products in microcentrifuge at 16k×*g* for 20–30 min.
14. Vacuum or air-dry the pellet. Expect good signals for the next step when the radioactive signal is detectable by the Geiger counter.

3.4 Thin-Layer Chromatography Reveals the Modification Status

1. Resuspend and dissolve the alcohol-precipitated RNA pellet with 3 μl nuclease P1 mixture (0.33 U/μl nuclease P1 in 30 mM sodium acetate/acetic acid, pH 4.8).
2. Incubate at 37 °C for 2 h to allow complete digestion.

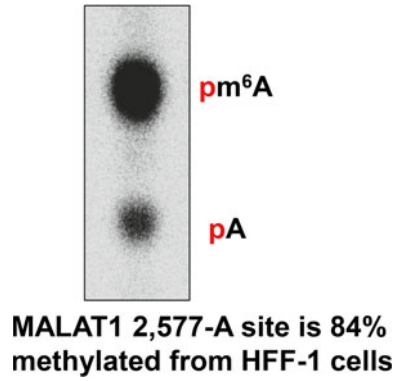

Fig. 3 TLC result showing the modification status of the 2577-A site on the MALAT1 lncRNA (NR_002819) from HFF-1 cells

3. Spot the reaction mix 0.5–1 μl at a time on a cellulose TLC plate (20 × 20 cm; Merck) as previously described. If multiple spotting is needed, wait for the TLC plate to dry completely before spotting the next aliquot and ensure spotting at the same position of the TLC plate.
4. Develop the TLC plate in a tank with running buffer [isopropanol:HCl:water (70:15:15, v/v/v), 100 ml]. This process takes ~14 h.

Day 3

5. After that, dry the TLC plate at room temperature for 1 h, wrap the plate in plastic film and expose it to a blanked phosphorimager screen. The exposure time varies from 1 to 20 h, depending on the strength of radioactive signals on the TLC plate.
6. Visualize and quantify the TLC result through the phosphorimager to get the modification status of the target nucleotide (Fig. 3). Please *see* **Notes 1–7**.

4 Notes

1. RNA modifications in abundant RNA, such as ribosomal RNA, small nuclear RNA, etc., normally generate very strong SCARLET signal.
2. RNA modifications in abundant mRNA/lncRNA tend to have stronger SCARLET signal and lower background noise signal on the TLC plate.
3. It is strongly advisable to use specific, synthetic oligos as reaction and TLC controls. Synthetic oligos normally ensure significant signal of m^{6}A on the electrophoresis gel and TLC

result, providing valuable information whether every step is done correctly, all reagents and enzymes are active, and the correct position of the bands or dots containing the target nucleotides.

4. The m^{6}A RNA modification fraction on mRNA/lncRNA normally ranges between 5 and 88 % [10]. We arbitrarily set the detection threshold at 5 %, so modification signal with a fraction less than 5 % is considered background noise or unmodified.
5. Since each SCARLET experiment works on only one candidate site, careful works on selecting candidate sites are needed to ensure successful detection of m^{6}A RNA modification. Potential candidate m^{6}A sites are evaluated through previous m^{6}A/MeRIP results, RNA abundance through previous RNA-seq data or else, RRACH consensus motif [15, 16], structural motif embedded, species-conserved level, and so on. More detailed selection process was described in ref. 10.
6. We normally perform ~10 SCARLET experiments in parallel to examine 10 potential m^{6}A sites at one time.
7. The modification fraction of m^{6}A located in structured RNAs can be underestimated by SCARLET, due to the inefficient hybridization with chimeric oligos [17].

Acknowledgments

We thank Dr. Q. Dai for assistance with chemical synthesis, and Dr. M. Parisien for bioinformatic analysis to choose potential target nucleotide sites from mRNA/lncRNA. This work was supported by a NIH grant (GM088599 to T.P. and C.H.).

References

1. Bokar JA (2005) The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In: Grosjean H (ed) Fine-tuning of RNA functions by modification and editing. Berlin, Springer, pp 141–178
2. Liu N, Pan T (2014) RNA epigenetics. Am J Transl Res 165:28–35
3. Pan T (2013) N^6-methyl-adenosine modification in messenger and long non-coding RNA. Trends Biochem Sci 38:204–209
4. Horowitz S, Horowitz A, Nilsen TW, Munns TW, Rottman FM (1984) Mapping of *N6-methyladenosine* residues in bovine prolactin mRNA. Proc Natl Acad Sci U S A 81: 5667–5671
5. Dai Q, Fong R, Saikia M et al (2007) Identification of recognition residues for ligation-based detection and quantitation of pseudouridine and N^6-methyladenosine. Nucleic Acids Res 35:6322–6329
6. Fu Y, Dominissini D, Rechavi G et al (2014) Gene expression regulation mediated through reversible m^{6}A RNA methylation. Nat Rev Genet 15:293–306
7. Dominissini D, Moshitch-Moshkovitz S, Schwartz S et al (2012) Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. Nature 485:201–206
8. Meyer KD, Saletore Y, Zumbo P et al (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 149:1635–1646
9. Chen K, Lu Z, Wang X et al (2014) High-resolution N^6-methyladenosine (m^{6}A) map using

photo-crosslinking-assisted m A sequencing. Angew Chem. doi:10.1002/anie.201410647

10. Liu N, Parisien M, Dai Q et al (2013) Probing N^6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. RNA 19:1848–1856
11. Zhao X, Yu YT (2004) Detection and quantitation of RNA base modifications. RNA 10:996–1002
12. Maroney PA, Chamnongpol S, Souret F et al (2008) Direct detection of small RNAs using splinted ligation. Nat Protoc 3:279–287
13. Wu G, Xiao M, Yang C et al (2011) U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. EMBO J 30:79–89
14. Yu YT, Shu MD, Steitz JA (1997) A new method for detecting sites of 2'-*O*-methylation in RNA molecules. RNA 3:324–331
15. Harper JE, Miceli SM, Roberts R et al (1990) Sequence specificity of the human mRNA *N6-adenosine* methylase in vitro. Nucleic Acids Res 18:5735–5741
16. Liu J, Yue Y, Han D et al (2014) A METTL3-METTL14 complex mediates mammalian nuclear RNA N^6-adenosine methylation. Nat Chem Biol 10:93–95
17. Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T (2015) N^6-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature 518:560–564

Part VI

Other Aspects of PTR

Chapter 18

Genome-Wide Identification of Alternative Polyadenylation Events Using 3′T-Fill

Stefan Wilkening, Vicent Pelechano, and Lars M. Steinmetz

Abstract

Due to the increasing appreciation of the impact of alternative polyadenylation on cellular biology, our straightforward, scalable method is of interest to any researcher studying eukaryotic transcription. In addition to high quality gene expression measurements, it precisely maps poly(A) sites and thereby permits the distinction between differential 3′UTR isoforms. As sequencing through long homopolymer stretches is not possible on the Illumina platform, we developed a method that fills up the poly(A) stretch with dTTPs before the sequencing reaction starts.

Key words Alternative polyadenylation, RNA-Seq, Mapping of poly(A) sites, Dark T-fill, Expression quantification

1 Introduction

The length of the 3′UTR and the regulatory sequences (like micro RNA-binding sites) within it determine the posttranscriptional fate of the mRNA molecules [1, 2]. Therefore, the identification of alternative poly(A) sites is of great interest for the study of gene expression. As shortening of 3′UTRs appears to be common in development and cancer, our technique is of special interest for researchers in these fields. Standard RNA-Seq methods are not suited to detect the exact poly(A) site as this part of the molecule is rarely sequenced and mapped. Sequencing through the poly(A) stretch [3] or following poly(A) tail shortening [4] normally leads to reduced sequencing quality. An explanation for that might be a variation in poly(A) stretches generated by polymerase slippage during the clustering [5]. Clustering is the process by which, prior to Illumina sequencing, individual library molecules are amplified through a solid-phase bridging PCR [6]. A cluster of molecules with length variations in the poly(A) stretch will result in desynchronized sequencing, as some molecules within one cluster will still incorporate a T, while others would already start incorporating

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_18,

the first 3′UTR base. Attempts to avoid reading through the homopolymer using a custom sequencing primer with a T-stretch [7, 8] do not completely solve this problem as they still lead to desynchronized sequencing start points and decreased base call quality. To address that problem, we developed a method (3′T-fill) that evades this problem by filling in the homopolymer stretch (independently of its length) before sequencing [5] (Fig. 1). Our 3′T-fill method also avoids sensitive RNA handling and the use of restriction enzymes prone to biased RNA quantification. This method also quantifies mRNA expression in a length-independent manner (unlike standard RNA-Seq methods), as only one read is produced per mRNA molecule. Filling of the poly(A) stretch with unlabeled dTTPs inside the Illumina clustering machine allows sequencing to start directly after the poly(A) tail into the 3′UTR. This is a simple and scalable method that requires less than 2 days for one person to produce dozens of sequencing libraries using standard laboratory equipment. We have successfully applied 3′T-fill to a wide variety of samples from yeast to humans, and even quantified specific 3′ mRNA isoforms in immunoprecipitated RNA [9], demonstrating its broad applicability. Since its development we have further improved this approach optimizing the dark T-fill reaction and making use of the dedicated indexing read from Illumina.

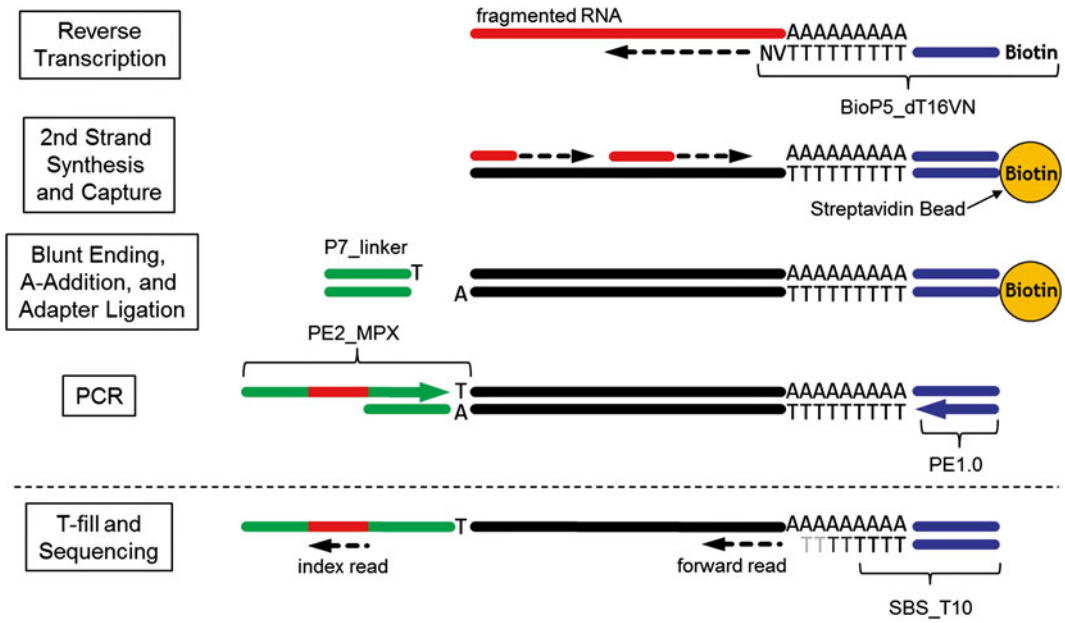

Fig. 1 Overview of the 3′T-fill protocol. Fragmented RNA is reverse transcribed using a oligo(dT) primer coupled to a biotinylated Illumina adapter. After second-strand synthesis, fragments are captured on streptavidin beads and part of the second Illumina adapter is ligated to the fragment. The PCR is performed with PE1.0 in combination with PE2_MPX baring a specific index per amplified sample. The A-stretch is filled in with complementary, unlabeled dTTPs on the cluster station. Sequencing starts directly at the end of the 3′ UTR and with a second primer the index (6–10 nucleotides) is sequenced

2 Materials

To avoid degradation of RNA by RNases, gloves should be worn and pipettes and working surfaces should be cleaned with an RNase-removing detergent. RNA should be kept on ice between steps. Prepare all solutions using ultrapure RNase-free water. A thermocycler with a heated lid is used for all incubations above room temperature.

2.1 Oligonucleotides

BioP5_dT16VN	[Biotin]ATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATCT TTTTTTTTTTTTTTTVN
P7_linker_for	GTGACTGGAGTTCAGACGTGTGCTCTT CCGATC*T
P7_linker_rev	[Phos]GATCGGAAGAGCACACGTCTGAACTC CAGTCAC[AmC7]
PE1.0	ATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTTCCGATC*T
PE2_MPX	CAAGCAGAAGACGGCATACGAGAT-Index- GTGACTGGAGTTCAGACGTGTGCTC TTCCGATC*T
SBS_T10	ACACTCTTTCCCTACACGACGCTCTTCCGA TCTTTTTTTTTT

Oligonucleotide sequences are based on Illuminas TruSeq® Small RNA Sample Prep Kit. Abbreviations: [Phos] = 5′ phosphorylation, * = s-linkage, [AmC7] = 3′Amine C7, -Index- = 6–10 nucleotides for multiplexing.

2.2 Reverse Transcription

1. 5× RNA fragmentation buffer: 200 mM Tris-acetate (pH 8.1), 500 mM KOAc, 150 mM MgOAc.
2. Polyadenylated control RNA: Transcribe in vitro from plasmids containing a T3 promoter and a DNA encoded poly(A) tail. The used cDNA should not be present in the organism of interest and be followed by 40–80 As. Isolate, quantify, and dilute this control RNA to a 100× solution corresponding to 50 ng/μL for 1–2 kb transcript controls [5].
3. Elution buffer (EB): 10 mM Tris–HCl (pH 8).
4. Actinomycin D solution: Dissolve 1.25 mg of actinomycin in 1 mL of water.
5. SuperScript II Reverse Transcriptase kit: 5× First-strand buffer, 100 mM DTT (Life Technologies).
6. RNasin Plus (Promega).

2.3 Second Strand Synthesis and Capture

1. DNA polymerase I and 10× buffer (Fermentas).
2. RNaseH (NEB).
3. Dynabeads M-280 Streptavidin (Life Technologies).
4. 2× bind and wash buffer (2× B&W): 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl. From this, prepare a 1× bind and wash buffer (1× B&W) by a 1:1 dilution with water.

2.4 Adapter Ligation

1. P7_linker: Mix the oligonucleotides *P7_linker_for* and *P7_linker_rev* (*see* Oligonucleotides 2.1) at 2.5 μM in the presence of 40 mM Tris–HCl (pH 8.0) and 50 mM NaCl. Incubate the sample for 5 min at 95 °C and let it slowly cool (−0.1 °C/s) to 65 °C. Incubate the sample for 5 min at 65 °C and let it slowly cool (−0.1 °C/s) to 4 °C. Store aliquots of the adapters at −20 °C. Thaw the linkers on ice to prevent denaturation or degradation. Annealing efficiency can be checked using a 2 % agarose gel comparing the migration of the annealed linkers to the individual oligonucleotides.
2. NEBNext DNA Sample Prep Master Mix Set 1 (NEB): 10× End repair buffer, end repair enzyme mix. Klenow Fragment ($3' \rightarrow 5'$ exo$^-$).
3. Ampure XP beads (Beckman Coulter).
4. T4 DNA ligase (2000 U/μL), 2× Quick Ligation buffer (NEB).
5. DA tailing buffer: Supplement 10× NEB buffer 2 with 0.2 mM dATP (NEB).

2.5 Enrichment PCR and Quality Control

1. Phusion® High-Fidelity DNA Polymerase (2 U/μL), Phusion HF Buffer (NEB).
2. Qubit (Life Technologies).
3. Bioanalyzer or TapeStation (Agilent).

2.6 Clustering

1. Taq Polymerase E (5 U/μL, Genaxxon).

3 Methods

3.1 Fragment Total RNA

1. Mix 10 μg of DNA-free total RNA (*see* **Note 1**) with 4 μL of 5× RNA fragmentation buffer in 0.2 mL PCR strips (*see* **Note 2**) or a 96-well plate, and optionally, add 4 μL of in vitro-transcribed control RNA for quality control. Adjust the total volume to 20 μL with RNase-free water and incubate the sample at 80 °C for 5 min. Transfer the sample to ice immediately.
2. Clean up with 1.5× Ampure XP beads (add 30 μL to the 20 μL fragmented RNA), and elute in 12.8 μL EB.

3.2 Reverse Transcription

1. Mix 11.2 µL of the fragmented RNA with 1 µL *P5_dT16VN* (1 µM) and 1 µL 10 mM dNTP Mix. If less starting material is used, the amount of oligo dT primer should be adjusted accordingly (*see* **Note 1**).
2. Incubate the sample for 5 min at 65 °C to disrupt secondary structures and place on ice.
3. Add 4 µL 5× first-strand buffer, 2 µL DTT, 0.32 µL actinomycin D, and 0.5 µL RNasin Plus to each sample. Mix tubes and put in a thermocycler at 42 °C.
4. To prevent mispriming of the anchored oligo(dT) primer *BioP5_dT16VN* wait for 2 min before adding 0.5 µL Superscript II to each tube without taking the samples out of the thermocycler.
5. Incubate the samples at 42 °C for 50 min, and then inactivate the enzyme by incubation at 72 °C for 15 min.
6. Clean up with 1.5× Ampure XP beads and elute in 40 µL EB. This will clean the excess of unused biotinylated oligo.

3.3 Second-Strand Synthesis

1. Mix 40 µL cDNA from the previous step with 5 µL 10× DNA polymerase I buffer and 2.5 µL dNTPs (10 mM) on ice.
2. Add 0.5 µL RNaseH and 2 µL DNA polymerase I.
3. Incubate the tubes at 16 °C for 2.5 h.
4. Cleanup with 0.9× Ampure XP beads and elute in 20 µL EB. This will remove the enzymes and short DNA fragments.

3.4 Capture of the Biotinylated 3′ Terminal cDNA Fragments

1. Prepare the 20 µL Dynabeads M-280 Streptavidin for each sample (good for binding 4 pmol of biotinylated oligo(dT)) according to the manufacturer's instructions.
2. Transfer the tubes to a magnetic stand and remove the supernatant (*see* **Note 3**).
3. Take the tubes from the magnetic stand and resuspend the beads in 200 µL 1× B&W.
4. Transfer the tubes to a magnetic stand and remove the supernatant.
5. Repeat the wash with 200 µL 1× B&W.
6. Transfer the tubes to a magnetic stand and remove the supernatant.
7. Resuspend the beads in 20 µL 2× B&W.
8. Add 20 µL of cDNA to the prepared beads. Mix and put on a rotator wheel at room temperature for 15 min to allow the capture of the ds-cDNA.
9. Using a magnet stand, wash two times with 200 µL 1× B&W and one time with 200 µL EB and resuspend in 21 µL EB.

3.5 Adapter Ligation

1. Mix the sample containing the beads from the previous step with 2.5 μL 10× end repair buffer and 1.25 end repair enzyme mix.
2. Incubate the tubes for 30 min at 20 °C.
3. Wash the beads twice with 200 μL 1× B&W and once with 200 μL EB and resuspended in 21 μL EB.
4. Mix the sample with 2.5 μL of dA tailing buffer and 1.5 μL Klenow Fragment (3′ → 5′ exo⁻).
5. Incubate the tubes for 30 min at 37 °C.
6. Transfer the tubes to a magnetic stand (*see* **Note 3**). Wash the beads twice with 200 μL 1× B&W and once with 200 μL EB and resuspended in 8 μL EB.
7. Thaw the double-stranded P7_linker on ice.
8. Mix the sample containing the beads from the previous step with 12.5 μL 2× Quick Ligation buffer, 2 μL P7_linker and 2.5 μL T4 DNA ligase. If less than 10 μg of starting material is used (*see* **Note 1**), the adapter concentration should be adapted accordingly to avoid adapter self-ligation.
9. Incubate the samples for 15–30 min at 20 °C. If using a normal ligase incubate 1–3 h at 16 °C.
10. Transfer the tubes to a magnetic stand and remove the supernatant.
11. Wash 4 times with 200 μL 1× B&W and once with 200 μL EB and resuspend in 50 μL EB.

3.6 Enrichment PCR

1. Since an excess of input DNA inhibits PCR amplification, we use only part of the beads (usually ½) for the PCR amplification. The optimal proportion and cycle number should be tested for each setup individually.
2. Mix 24 μL beads with DNA with 25 μL 2× Phusion Master Mix with HF buffer and 0.5 μL of the primer *PE1.0* and *PE2_MPX* (10 μM) each. Use a different *PE2_MPX* for each sample you want to multiplex. Complexity of barcodes (different nucleotide at each position) should be maximized to avoid problems during Illumina base calling.
3. Incubate the sample in a thermocycler with the following program: 30 s at 98 °C, 18 cycles (10 s at 98 °C, 10 s at 65 °C and 10 s at 72 °C), and a final extension of 5 min at 72 °C.
4. Transfer the PCR product to a magnetic stand (*see* **Note 3**) and recover the 50 μL supernatant containing the PCR product.
5. The beads used for the PCR should be kept at 4 °C in case a re-amplification is needed.
6. Purify the eluted PCR product with 1.8× Ampure XP beads, and elute in 10 μL EB.

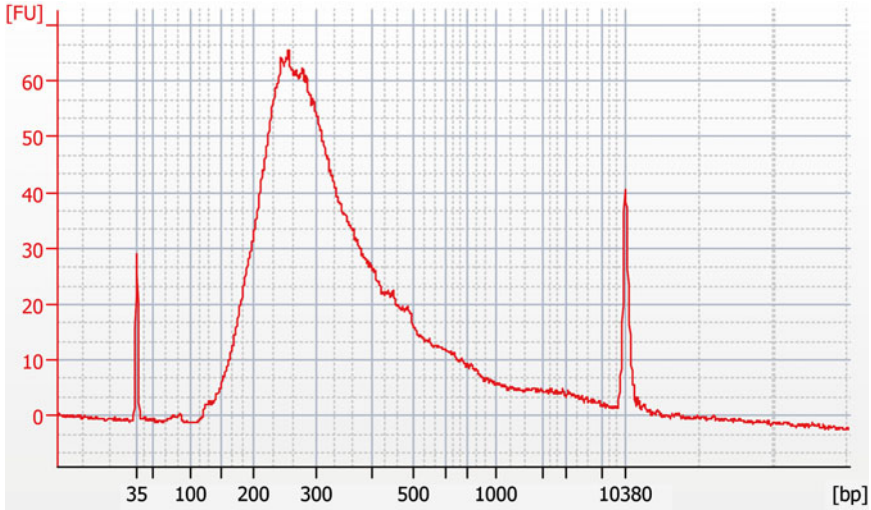

Fig. 2 Example for a size distribution of a 3′T-fill library analyzed on a Bioanalyzer (Agilent)

3.7 Quality Control

1. Measure the concentration by Qubit, it should be >10 ng/μL.
2. Determine the size distribution by loading 1 μL of the final sample in a Bioanalyzer or TapeStation. Primer dimer (100–120 bp should not be present). Library size around 250–350 bp is optimal (Fig. 2).
3. When preparing 3′T-fill library for the first time, we recommend cloning a library amplified with a taq polymerase (for A-addition) in *E. coli* using a TA cloning kit and to Sanger sequence some clones to confirm the correct construction.

3.8 Clustering and Sequencing

1. A key step to avoid problems caused by sequencing through the poly(A) tail is the exchange of the hybridization buffer (HT1) with 130 μL of the following buffer: 101 μL water, 20 μL 5× Phusion HF Buffer, 3 μL dTTPs (10 mM), 0.8 μL *SBS_T10* (100 μM), 3 μL Phusion® High-Fidelity DNA Polymerase, and 2 μL Taq Polymerase E (*see* **Note 4**).
2. In an Illumina HiSeq 2000 we normally get between 100 and 200 million clusters with a 50-cycle single-end run.

4 Notes

1. We normally use 10 μg total RNA as starting amount for this protocol, but started with as low as 500 ng total RNA without significant decrease in quality.
2. To process multiple samples simultaneously, it is advisable to use 0.2 mL PCR strips or 96-well plates for library preparation in combination with multichannel pipettes.

3. For the cleanup with AMPure XP beads follow the manufactures protocol (https://www.beckmancoulter.com/). For the purification in 0.2 mL PCR-strips we use a self-made magnetic stand as described in Wilkening et al. [10]. Throughout the protocol, different ratios of beads are used, in general, higher amounts of bead solution increase the capture of shorter DNA molecules [11].
4. We also obtained good results using the QuantSeq 3′mRNA kit from Lexogen that adapted the 3′T-fill approach for library preparation.

Acknowledgement

We would like to thank Manu M. Tekkedil, Aino I. Järvelin, and Vladimir Benes for their help in preparing the original protocol and EMBL Genomics Core Facility for technical support. This work was supported by grants from the University of Luxembourg-Institute for Systems Biology Program and Deutsche Forschungsgemeinschaft (to L.M.S.) and an EMBO fellowship (to V.P.).

References

1. Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Mol Cell 43:853–866
2. Lutz CS, Moreira A (2011) Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip Rev RNA 2:22–31
3. Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA 16:1256–1267
4. Jan CH, Friedman RC, Ruby JG, Bartel DP (2011) Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. Nature 469:97–101
5. Wilkening S, Pelechano V, Jarvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM (2013) An efficient method for genome-wide polyadenylation site mapping and RNA quantification. Nucleic Acids Res 41(5):e65
6. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11:31–46
7. Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17:761–772
8. Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res 22:1173–1183
9. Gupta I, Clauder-Munster S, Klaus B, Jarvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM (2014) Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. Mol Syst Biol 10:719
10. Wilkening S, Tekkedil MM, Lin G, Fritsch ES, Wei W, Gagneur J, Lazinski DW, Camilli A, Steinmetz LM (2013) Genotyping 1000 yeast strains by next-generation sequencing. BMC Genomics 14:90
11. Lundin S, Stranneheim H, Pettersson E, Klevebring D, Lundeberg J (2010) Increased throughput by parallelization of library preparation for massive sequencing. PLoS One 5:e10029

Chapter 19

Genome-Wide Profiling of Alternative Translation Initiation Sites

Xiangwei Gao, Ji Wan, and Shu-Bing Qian

Abstract

Regulation of translation initiation is a central control point in protein synthesis. Variations of start codon selection contribute to protein diversity and complexity. Systemic mapping of start codon positions and precise measurement of the corresponding initiation rate would transform our understanding of translational control. Here we describe a ribosome profiling approach that enables identification of translation initiation sites on a genome-wide scale. By capturing initiating ribosomes using lactimidomycin, this approach permits qualitative and quantitative analysis of alternative translation initiation.

Key words Ribosome profiling, Translation, Initiation, Start codon, Genome-wide, Deep sequencing

1 Introduction

Translation initiation entails the ordered assembly of translation-competent ribosomes with initiator tRNA basepaired to the mRNA start codon at the ribosomal P-site [1]. In eukaryotes, the translation start codon is generally identified by the scanning mechanism [2]. It is believed that the first AUG codon encountered by the 48S preinitiation complex serves as the translation start codon. However, an increasing body of evidence suggests that eukaryotic ribosomes can recognize several alternative translation initiation sites (TISs) [3, 4]. It has been estimated that about 50 % of mammalian transcripts contain at least one upstream open reading frame (uORF) [5, 6]. To confound the matter further, many non-AUG triplets have been reported to act as alternative TISs for translation initiating [7]. Since there is no reliable way to predict non-AUG codons as potential initiators from *in silico* sequence analysis, there is a pressing need to develop experimental approaches for genome-wide TIS identification.

Ribosome profiling, based on deep sequencing of ribosome-protected mRNA fragments (RPF), has proven to be powerful in defining ribosome positions on the entire transcriptome [8].

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_19,

However, the standard ribosome profiling captures all the ribosomes engaged on the mRNA. Therefore, it is not optimized for identifying TIS positions. Several strategies have been developed in order to achieve efficient capture of initiating ribosomes. Translation inhibitors like harringtonine act on the first round of peptide bond formation and potentially stall the initiating ribosomes [9]. Puromycin dissociates the ribosome into subunits but exhibits higher sensitivity towards elongating ribosomes than initiating ribosomes [10]. However, both compounds do not seem to arrest the ribosome at the start codon in a definitive manner. To improve the resolution of TIS mapping, we took advantage of a translation inhibitor lactimidomycin (LTM) that acts preferentially on the initiating ribosomes [11]. Two different approaches were developed to effectively capture the initiating ribosomes from mammalian cells (Fig. 1). The first approach, global translation initiation sequencing (GTI-seq), permits high-resolution mapping of TIS positions via LTM pretreatment [12]. The second approach, quantitative translation initiation sequencing (QTI-seq), enables

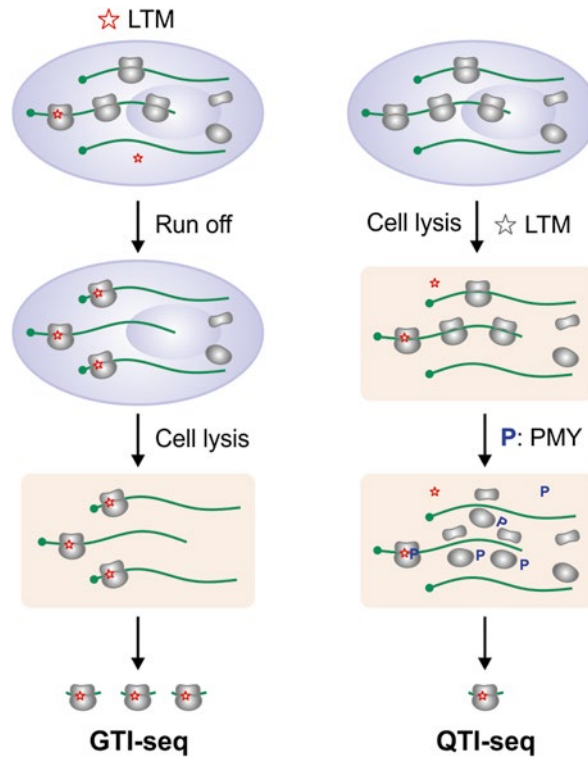

Fig. 1 Schematic of GTI-seq (*left*) and QTI-seq (*right*). In GTI-seq, cells are pretreated with LTM to immobilize the initiating ribosomes followed by an incubation to dissociate the elongating ribosomes. In QTI-seq, LTM directly captures the initiating ribosomes from the lysates and relies on puromycin (PMY) to dissociate the elongating ribosomes

enrichment of initiating ribosomes from whole cell lysates [13]. Along with detailed methods, both the advantages and the disadvantages of these approaches are discussed in this chapter.

2 Materials

2.1 Preparation of Cell Lysates for GTI-seq

1. Growth medium: Dulbecco's modified Eagle medium (DMEM) supplemented with 10 % fetal bovine serum (FBS).
2. Phosphate-buffered saline (PBS).
3. 10 cm petri dishes.
4. Lactimidomycin: Dissolve in DMSO to the concentration of 50 mM and store it at −80 °C.
5. Cycloheximide: Dissolve in DMSO to the concentration of 100 mg/mL and store it at −80°C.
6. GTI-seq lysis buffer: 20 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$, 100 μg/mL cycloheximide, 2 % Triton X-100.

2.2 Preparation of Cell Lysates for QTI-seq

In addition to the materials for GTI-seq, other materials are listed below:

1. Lysing Matrix D.
2. QTI-seq buffer: 20 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$.
3. Puromycin: Dissolve in ddH_2O to the concentration of 20 μg/mL and store it at −20 °C.
4. Creatine phosphate: Dissolve in nuclease-free H_2O to the concentration of 1 M and store it at −20 °C.
5. Spermidine: Dissolve it in nuclease-free H_2O to the concentration of 10 mM and store it at −20 °C.
6. Creatine phosphokinase: Dissolve in nuclease-free H_2O to the concentration of 10 mg/mL and store it at −20 °C.
7. ATP solution (10 mM).

2.3 Sucrose Gradient Sedimentation

1. Polysome buffer: 10 mM Hepes, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$.
2. Sucrose solution: First make the 60 % (wt/vol) sucrose solution using polysome buffer. Mix 1 volume of 60 % sucrose and 3 volumes of polysome buffer to get 15 % sucrose. Mix 3 volumes of 60 % sucrose and 1 volume of polysome buffer to get 45 % sucrose.
3. SW41 ultracentrifuge tubes.
4. Sucrose gradient maker.
5. Automated fractionation system.

2.4 RNase I Digestion and RPF Extraction

1. RNase I.
2. TRIzol LS reagent.
3. T4 Polynucleotide Kinase.
4. SUPERase_In.
5. 2× Novex TBE-urea sample buffer.
6. 10× Novex TBE-urea gel running buffer.
7. Novex 15 % denaturing polyacrylamide TBE-urea gel.
8. SYBR Gold: Dilute to 1:10,000 in TAE buffer and store it at −20 °C.
9. Spin-X column.
10. Glycogen (5 mg/mL).
11. 3 M NaOAc (pH 5.2).
12. RNA gel elution buffer: 300 mM NaOAc (pH 5.5), 1 mM EDTA, and 0.1 U/μL SUPERase_In. Store the buffer at −20 °C.

2.5 cDNA Library Construction

1. *E. coli* Poly(A) polymerase.
2. SuperScript III Reverse Transcriptase.
3. RNaseOUT Recombinant Ribonuclease Inhibitor:.
4. Reverse transcription primers: 5′-pGATCGTCGGACTGTA GAACTCTØCAAGC AGAAGACGGCATACGATTTTTTTT TTTTTTTTTTTTVN-3′. The initial p indicates 5′ phosphorylation, Ø indicates the abasic dSpacer furan, and V and N indicate degenerate nucleotides.
5. Novex 10 % denaturing polyacrylamide TBE-urea gel.
6. DNA gel elution buffer: 300 mM NaCl and 1 mM EDTA. Prepare in advance and store at 4 °C.

2.6 DNA Circularization and Re-linearization

1. Circularization kit.
2. Ape I.

2.7 DNA Amplification and Deep Sequencing

1. Phusion High-Fidelity DNA Polymerase.
2. Oligo primers for PCR: sense 5′-CAAGCAGAAGACGGC ATA-3′; antisense: 5′-AATGATACGGCGACCACCGACAG GTTCAGAGTTCTACAGTCCGACG-3′ to generate DNA species containing Illumina cluster generation sequences on each end and a sequencing primer-binding site.
3. Novex 8 % denaturing polyacrylamide TBE-urea gel.
4. 5× Novex TBE Running Buffer.

3 Methods

3.1 Preparation of Cell Lysates for GTI-seq

In this approach, human embryo kidney (HEK) 293 cells are pretreated with LTM for 30 min to allow the elongating ribosomes to run off. Different cell lines have varied LTM sensitivity and may require pilot experiments to determine both the LTM concentration and the incubation period (*see* **Note 1**). The following experimental procedure is based on HEK293 cells.

1. Grow HEK293 cells in growth medium in a 5 % CO_2 cell culture incubator with 95 % humidity. Split HEK293 cells the day before the treatment to at least four 10 cm petri dishes. Adjusting the splitting ratio so that the cells are at approximately 80 % confluence on the day of experiment.
2. Replace the medium with 5 mL fresh, pre-warmed medium containing 50 μM LTM followed by incubation for 30 min (*see* **Note 2**).
3. Aspirate medium from each dish and immediately cool the dishes on ice. Gently wash the cells with 5 mL of ice-cold PBS containing 100 μg/mL of cycloheximide.
4. Aspirate the PBS thoroughly from all the petri dishes and place them on ice.
5. Add 400 μL of ice-cold lysis buffer to one petri dish (*see* **Note 3**).
6. Detach the cells by scraping the entire petri dish and pipetting up and down several times to lyse the cells.
7. Transfer the lysates from the first petri dish to the second dish and lyse the cells. Repeat **step 6** until all the petri dishes are done (*see* **Note 4**).
8. Transfer the cell lysates from the last petri dish to a 1.6 mL Eppendorf tube and place the tube on ice.
9. Remove the debris from the cell lysates by centrifugation for 10 min at 13,000 × *g* at 4 °C. Transfer the soluble supernatant to a new 1.6 mL Eppendorf tube and store the tube on ice.

3.2 Preparation of Cell Lysates for QTI-seq

In this approach, human embryo kidney (HEK) 293 cells are used to enrich initiating ribosomes from the cell lysates. Sequential treatment of LTM and puromycin is conducted on the cell lysates. The detailed experimental procedure is described below.

1. Grow HEK293 cells in growth medium in a 5 % CO_2 cell culture incubator with 95 % humidity. Split HEK293 cells the day before the treatment to at least four 10 cm petri dishes. Adjusting the splitting ratio so that the cells are at approximately 80 % confluence on the day of experiment.
2. Aspirate medium from each dish and immediately cool the dishes on ice. Gently wash the cells with 5 mL of ice-cold PBS.

3. Aspirate the PBS thoroughly from all the petri dishes and place them on ice.
4. Add 400 μL of ice-cold QTI-seq buffer containing 5 μM LTM to the first petri dish (*see* **Note 4**).
5. Detach the cells by scraping the entire petri dish and pipetting up and down several times to suspend all the cells.
6. Transfer the buffer from the first petri dish to the second dish to detach the cells. Repeat **step 5** until all the petri dishes are done.
7. Transfer the collected cells from all the petri dishes to a 2 mL Eppendorf tube containing Lysing Matrix-D and place the tube on ice.
8. Lyse cells by vortexing 20 s for six times with a 40-s interval on ice.
9. Remove the debris from the cell lysates by centrifugation for 10 min at 13,000 × *g* at 4 °C. Transfer the soluble supernatant to a new 1.6 mL Eppendorf tube and store the tube on ice.
10. Supplement the following reagents into the cell lysates: 10 mM creatine phosphate, 0.1 mM spermidine, 40 μg/mL creatine phosphokinase, 0.8 mM ATP, and 25 μM of puromycin. Incubate the mixture at 35 °C for 15 min.

3.3 Sucrose Gradient Sedimentation

For both GTI-seq and QTI-seq, it is important to examine whether the dissociation of elongating ribosomes is complete. Sucrose gradient-based polysome profiling is commonly used for this purpose. As shown in Fig. 2, a complete dissociation of elongating ribosomes results in disassembly of polysome fractions with a corresponding increase of monosome.

1. Freshly make 12 mL sucrose density gradients in each SW41 ultracentrifuge tube. Use the BioComp Gradient Master (BioComp Inc.) to generate a 15–45 % (wt/vol) gradient. In brief, first fill the tube with 6 mL of 15 % sucrose solution

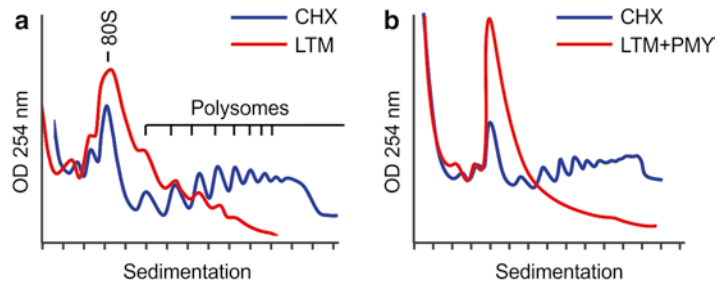

Fig. 2 Polysome profiling to examine the dissociation of elongating ribosomes during GTI-seq (**a**) and QTI-seq (**b**)

and then add 6 mL of 45 % sucrose solution to the bottom of the tube using needle and syringe. Follow the manufacturer's instruction to mix the two layers of sucrose solution.

2. Gently load 600 μL of cell lysate onto the top surface of the sucrose gradients, followed by centrifugation for 100 min at 38,000 × *g*, 4 °C, in an SW41 rotor (*see* **Note 5**).
3. Fractionate the separated samples using an automated fractionation system (Isco) with continuous monitoring of OD_{254} values. Fractions are collected into 1.6 mL Eppendorf tubes at 1.5 mL/min with 0.5 min per tube. The collected fractions can be either stored at −80 °C or immediately used for RNase I digestion.

3.4 RNase I Digestion and RPF Extraction

Nuclease digestion removes regions of mRNAs not protected by the ribosome, converting polysome into monosome carrying the RPF. The nuclease choice and the digestion condition need to be carefully considered to avoid incomplete digestion or over-digestion. Following nuclease digestion, the RPFs are resolved on a gel and collected via a size selection. The typical RPF size in mammalian cells is around 28 nt (Fig. 3).

1. Take 10 μL from each fraction starting from the monosome and mix thoroughly to get a total of 300 μL ribosome fractions in a 1.6 mL Eppendorf tube.
2. Add *E. coli* RNase I into the mixed ribosome fractions (750 U per 100 A_{260} units). Incubate the mixture at 4 °C for 1 h to convert the polysome into monosome (*see* **Note 6**).
3. Add 3 volume of TRIzol LS reagent to the digested ribosome samples. Extract total RNA according to the manufacturer's instructions. Dissolve the total RNA in 11 μL of RNase-free H_2O.
4. Dephosphorylate the purified RNA samples in a 15-μL reaction mixture by adding 1.5 μL 10× T4 polynucleotide kinase buffer, 0.5 μL RNase inhibitor SUPERase_In (20 U/μL), and 2 μL T4 polynucleotide kinase (10 U/μL). Incubate the reaction mixture at 37 °C for 2 h, followed by heat-inactivation at 65 °C for 20 min.
5. Mix the dephosphorylated sample with 15 μL 2× Novex TBE-Urea sample buffer (Invitrogen). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 15 % denaturing polyacrylamide TBE-urea gel (Invitrogen) and run at 160 V for 1.5 h.
6. Stain the gel with SYBR Gold to visualize the RNA fragments. Excise the gel bands corresponding to 28 nt RNA species (Fig. 3).

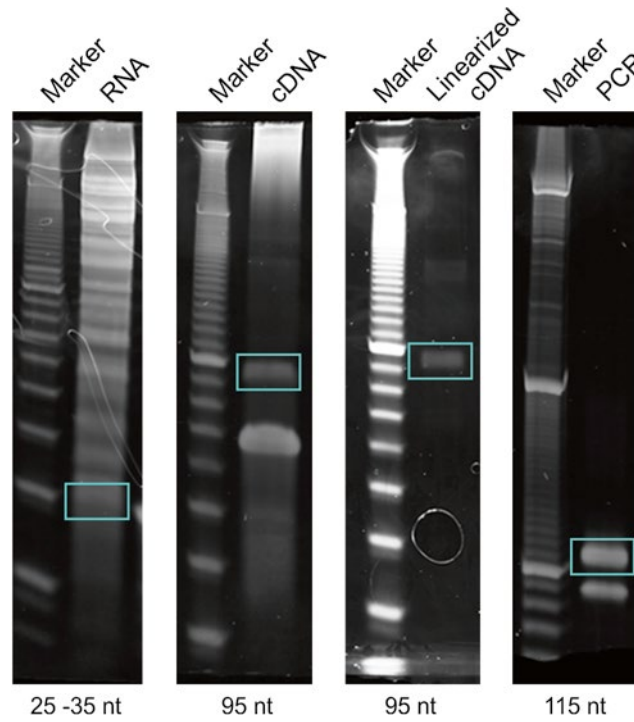

Fig. 3 Size selection during cDNA library construction. The ribosome-protected mRNA fragments, single stranded DNA, linearized DNA, and PCR products are *highlight in box*

7. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the RNA fragments by soaking the gel slices overnight in RNA gel elution buffer.
8. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
9. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
10. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
11. Aspirate the liquid and air-dry the pellet. Dissolve the RNA pellet in 5.2 μL of nuclease-free H_2O.

3.5 cDNA Library Construction

The entire procedure of cDNA library construction from purified RPF is similar to regular ribosome profiling [14], which includes poly(A) tailing and reverse transcription. The custom-designed primers may include barcodes for multiplex analysis (*see* **Note 7**).

1. Perform poly-(A) tailing reaction in a 8 μL reaction mixture by adding 0.8 μL poly-(A) polymerase buffer (10×), 1 μL ATP

(10 mM), 0.4 μL SUPERase_In, and 0.6 μL *E. coli* poly-(A) polymerase (5 U/μL). Incubate the reaction mixture at 37 °C for 45 min.

2. Mix the poly(A)-tailed RNA samples with 1 μL dNTP (10 mM) and 1 μL synthesized primer (25 mM). Incubate the reaction mixture at 65 °C for 5 min followed by ice incubation for 5 min.
3. For reverse transcription, set up a 20 μL reaction mixture by adding 2 μL 10× buffer, 4 μL $MgCl_2$ (25 mM), 2 μL DTT (100 mM), 1 μL RNaseOUT (40 U/μL), and 1 μL SuperScript III (200 U/μL). Incubate the reaction mixture at 50 °C for 50 min.
4. Mix the reverse transcription sample with 20 μL Novex TBE-Urea sample buffer (2×). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 10 % denaturing polyacrylamide TBE-urea gel and run at 160 V for 1.5 h.
5. Stain the gel with SYBR Gold to visualize the single-strand DNA. Excise the gel bands corresponding to the size of 95 nt (Fig. 3).
6. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single-strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
7. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
8. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
9. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
10. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 11.5 μL nuclease-free H_2O.

3.6 DNA Circularization and Re-linearization

1. Circularize the single-stranded DNA in 20 μL of reaction mixture by adding 2 μL CircLigase buffer (10×), 1 μL $MnCl_2$ (50 mM), 4 μL Betaine (5 M), and 1.5 μL CircLigase II (100 U/μL). Circularization is performed at 60 °C for 1.5 h followed by heat-inactivation at 80 °C for 10 min (*see* **Note 8**).
2. Purify the circularized DNA using ethanol precipitation mentioned above. Dissolve the DNA pellet in 8.25 μL of nuclear-free H_2O.
3. Re-linearize the circular single-stranded DNA in 10 μL of reaction mixture by adding 0.75 μL APE 1 (10 U/μL) and 1 μL Buffer 4 (NEB). The reaction is carried out at 37 °C for 1 h (*see* **Note 9**).

4. Mix the re-linearized DNA sample with 10 μL Novex TBE-Urea sample buffer (2×). Denature the sample in a heat block at 70 °C for 3 min. Load the samples on a Novex 10 % denaturing polyacrylamide TBE-urea gel (Invitrogen) and run at 160 V for 1.5 h.
5. Stain the gel with SYBR Gold to visualize the single-strand DNA. Excise the DNA gel band corresponding to the size of 95 nt (Fig. 3).
6. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single-strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
7. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
8. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3 M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.
9. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
10. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 10 μL nuclease-free H_2O.

3.7 DNA Amplification and Deep Sequencing

1. Amplify the single-stranded DNA template by setting up a 25 μL PCR reaction mixture containing 5 μL DNA template, 5 μL HF buffer (5×), 0.5 μL dNTP (10 mM), 1.25 μL each primer (10 mM), 0.5 μL Phusion polymerase (2 U/μL), and 11.5 μL distilled H_2O.
2. Perform PCR with an initial denaturation at 98 °C for 30 s, followed by 8–14 cycles of 10-s denaturation at 98 °C, 20-s annealing at 60 °C, and 10-s extension at 72 °C (*see* **Note 10**).
3. Separate PCR products on a nondenaturing 8 % polyacrylamide TBE gel by running at 160 V for 1 h.
4. Stain the gel with SYBR Gold to visualize the PCR product. Excise the DNA band corresponding to the size of 115 bp (Fig. 3)
5. Physically disrupt the gel by using centrifugation through the bottom holes of the tube. Extract the single strand DNA by soaking the gel slices overnight in DNA gel elution buffer.
6. Remove the gel debris using a Spin-X column by centrifugation at 12,000 × *g* for 2 min at room temperature.
7. Mix the elution with 4 μL glycogen, 40 μL NaOAc (3M, pH 5.2), and 900 μL ice-cold ethanol. Place the sample at −20 °C for 30 min.

8. Centrifuge the mixture at 14,000 × *g* for 15 min at 4 °C. Wash the pellet once with 70 % ethanol and centrifuge at 14,000 × *g* for 5 min at 4 °C.
9. Aspirate the liquid and air-dry the pellet. Dissolve the DNA pellet in 15 μL nuclease-free H_2O.
10. Quantify the PCR product by Agilent BioAnalyzer DNA 1000 assay.
11. Use approximately 3–5 pmol mixed DNA samples for cluster generation followed by sequencing using Illumina sequencing primer.

3.8 Sequencing Data Analysis

We conduct the sequencing data analysis using custom made R and Python scripts. The quality of TIS profiles can be evaluated by an aggregation plot showing all the RPF density across the transcriptome aligned at the annotated start codon. A high-quality TIS mapping is expected to see a significantly higher peak at the annotated start codon than other positions (Fig. 4a). A typical example of single gene TIS mapping is shown in Fig. 4b.

1. Categorize the raw reads of sequencing data into four groups according to barcode information (TG, AC, GA, and CT). After removing the 2-nt barcode, trim the reads by 10 nt from the 3′ end. The shortened reads are further trimmed by removing adenosine (A) stretch from the 3′ end with one mismatch allowed. Retain the processed reads for following analysis (*see* **Note 11**).
2. Map processed reads to transcriptome and hereafter unmapped reads to the corresponding genome using Tophat with parameters (--bowtie1 -p 10 --no-novel-juncs –G) [15]. Discard non-uniquely mapped reads to rule out ambiguity for the further analysis (*see* **Note 12**).

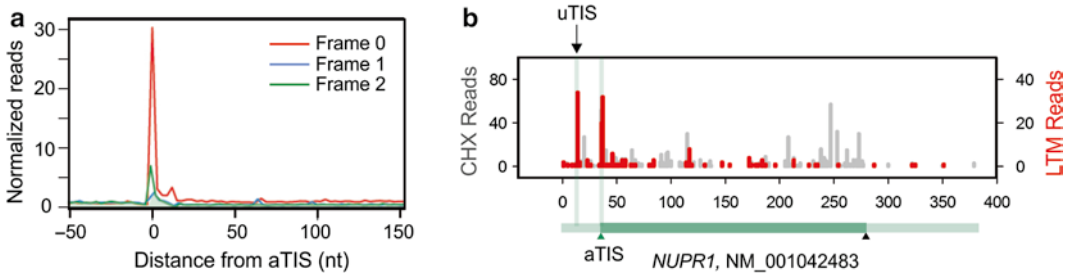

Fig. 4 Sequencing data analysis. (**a**) Aggregation plot of LTM-associated ribosome density in HEK293 cells captured by QTI-seq. Normalized RPF reads are averaged across the entire transcriptome and aligned at the annotated start codon. Different reading frames are separated and color coded. (**b**) An example of gene (NUPR1) with an alternative TIS captured by QTI-seq. The corresponding gene structure is shown below the *X*-axis with *green triangle* as the annotated start codon, and *black triangle* as the stop codon

3. Define the 13th position (12 nt offset from the 5′ end) of the uniquely mapped read as the ribosome "P-site". Calculate the P-site density on each position of individual mRNA transcript according to the NCBI Refseq annotation.
4. Apply Zero-Truncated Negative (ZTNB) model for identifying statistically significant P-site peak on the mRNA. Fit a global ZTNB model over all the non-empty P-sites of the entire transcriptome; and for each individual transcript with more than 50 distinct P-site positions, fit a local ZTNB model of the non-zero P-sites (*see* **Note 13**).
5. Decide putative "start codon" as a peak by the following criteria: (a) *p*-value thresholds (0.05 for global ZTNB model and 0.01 for local ZTNB model); (b) the peak height should be a local optimal number in a fixed window (−15, +15) (*see* **Note 14**).
6. To make pairwise comparison for TIS efficiency between different conditions, apply upper quartile (UQ) normalization to each predicted TIS sites based on the population of total read count on each individual mRNA. Calculate the fold change of each TIS to estimate the differential translational initiation signals between two experimental conditions. Calculate the abundance of translation initiation signal using a window centering the predicted TIS codon (−1, +4) because read length variation in the QTI-Seq data may cause offset problem in defining P-site (*see* **Note 15**).

4 Notes

1. The data quality of GTI-seq and QTI-seq appears to be dependent on cell lines. Some cells are more sensitive to LTM treatment than the others.
2. LTM is sensitive to light exposure. Shield the stock solution during the procedure.
3. To maximize the protein concentration of the lysates, it is recommended to use the minimum amount of the lysis buffer but sufficient to cover the petri dish.
4. Use the same lysis buffer for multiple dishes can minimize the total volume of the cell lysates.
5. Remove 600 μL of sucrose from the top of the gradient prior to the addition of the same amount of the sample. Do not disturb the gradient.
6. To avoid incomplete digestion or over-digestion, a pilot experiment is recommended to determine the digestion condition. A second sucrose gradient sedimentation of the sample is suggested to examine the status of digestion.

7. The sequence of bar code could be included in the custom designed primers for reverse transcription.
8. CircLigase I is also recommended for this step.
9. The step of re-linearization is optional. The circularization product can be used directly for PCR amplification.
10. It is recommended to minimize the cycles of PCR. A pilot PCR reaction with different cycles is suggested to determine the optimal cycles.
11. The size of RPF may affect the mapping accuracy. It is recommended to examine the size distribution of RPF as part of the quality control.
12. The non-uniquely mapped reads can be retained for certain analysis since they are true RPFs.
13. Global ZTNB model, which corresponds to a transcriptome-wide model trained on all the data, aims to suggest a peak height which is statically significant; while local ZTNB-model, which corresponds to a model only built on the data of individual gene, aims to compare peaks within the gene as well as control for the variation in gene expression levels.
14. It is suggested to compare different thresholds for optimal parameters to control false positive predictions.
15. Since the length range of RPFs is 25–35 nt, an offset of 1 or 2 nt is possible in TIS identification. A clustering step can be used to group neighboring putative TIS sites (within 1 or 2 nt) to avoid redundant peak calling.

Acknowledgements

This work was supported by National Institutes of Health (DP2 OD006449, R01AG042400), Ellison Medical Foundation (AG-NS-0605-09), and US Department of Defense (W81XWH-14-1-0068).

References

1. Hinnebusch AG (2014) The scanning mechanism of eukaryotic translation initiation. Annu Rev Biochem 83:779–812
2. Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34
3. Iacono M, Mignone F, Pesole G (2005) uAUG and uORFs in human and rodent 5'untranslated mRNAs. Gene 349:97–105
4. Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 20:8635–8642
5. Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A 106:7507–7512
6. Resch AM, Ogurtsov AY, Rogozin IB, Shabalina SA, Koonin EV (2009) Evolution of alternative and constitutive regions of mammalian 5'UTRs. BMC Genomics 10:162
7. Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S (2003) Generation of protein isoform diversity by alternative initi-

ation of translation at non-AUG codons. Biol Cell 95:169–178

8. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223
9. Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147:789–802
10. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. Genome Res 22:2208–2218
11. Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, Green R, Shen B, Liu JO (2010) Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol 6:209–217
12. Lee S, Liu B, Huang SX, Shen B, Qian SB (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci U S A 109:E2424–E2432
13. Gao X, Wan J, Liu B, Ma M, Shen B, Qian S (2014) Quantitative profiling of initiating ribosomes in vivo. Nat Methods 12(2):147–153
14. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc 7:1534–1550
15. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

Chapter 20

Genome-Wide Study of mRNA Isoform Half-Lives

Joseph V. Geisberg*, and Zarmik Moqtaderi*

Abstract

In eukaryotes, RNA polymerase II-driven transcription and processing results in the formation of numerous mRNA 3′ isoforms that for any given gene may differ from one another by as little as a single nucleotide. These 3′ isoforms can vary in physical properties that may affect their function and stability. Here, we outline a systematic framework to measure individual mRNA 3′ isoform half-lives on a genome-wide level in *S. cerevisiae*. Our approach utilizes the Anchor-Away system to sequester RNA polymerase II (Pol II) in the cytoplasm followed by direct single-molecule RNA sequencing to generate a highly detailed view of 3′ isoform stability under most physiological conditions without many of the adverse effects associated with commonly used alternative approaches.

Key words mRNA half-life, mRNA isoform, mRNA stability, Anchor-Away, *Saccharomyces cerevisiae* transcription, RNA polymerase II

1 Introduction

In recent years, it has become increasingly evident that transcription of individual genes by Pol II gives rise to a broad spectrum of mRNA isoforms. These 3′ isoforms can differ by as little as a single nucleotide or by more than 1 kb in sequence [1–3]. In *S. cerevisiae*, actively transcribed genes frequently exhibit >50 such isoforms [1]. More generally, broad 3′ isoform diversity has also been observed in other yeast species [4], Arabidopsis [2], as well as higher eukaryotes [5, 6].

Little is known about the functional importance, physical properties, or regulation of these alternate 3′ isoforms. However, several recent studies have demonstrated that alternate 3′ mRNA isoforms can play critical roles in a number of important biological processes. First, shorter 3′ isoforms that are missing microRNA (miRNA)-binding sites have been implicated in oncogenic transformation in mammalian cells [7], presumably due to the ability of these isoforms to evade destruction by the miRNA-based degradation pathways.

*Equal author contribution.

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_20,

Second, differential mRNA isoform utilization is critical for zebrafish development and might hinge on preferential, tissue-specific expression of long, U-rich-element-containing stable isoforms at the expense of shorter transcripts that are more rapidly turned over [8]. Third, alteration in poly(A) site usage has been reported to be important in a number of diseases [9], most recently in multiple forms of muscular dystrophy [10, 11]. While the molecular mechanisms by which alternate 3′ isoforms can alter biological pathways remain poorly understood, an emerging theme is that wholesale changes in isoform expression and stability often play critical roles in multiple biological processes.

In *S. cerevisiae*, the Anchor-Away method has been used with great success to conditionally and rapidly deplete numerous nuclear proteins [12, 13]. In Anchor-Away, the protein to be depleted is fused with the rapamycin-binding site from the FRB protein, while the "anchor" (typically RPS13, a highly abundant ribosomal protein) is tagged with the rapamycin-binding domain of FKBP12. In the process of ribosomal maturation, RPS13-FKBP12 chimera transits the nucleus and is exported to the cytoplasm. Addition of rapamycin results in the formation of a molecular complex between RPS13-FKBP and the target nuclear protein [13]. Once formed, this complex is rapidly exported into the cytoplasm, resulting in a near-complete depletion of the target protein from the nucleus. Two of Anchor-Away's strongest points are that it can be used with many different growth conditions and that it has a minimal impact on the stress response in yeast. Both of these characteristics make the Anchor-Away strategy ideally suited for the study of yeast mRNA half-lives.

Here, we provide the experimental framework necessary to measure and analyze *S. cerevisiae* mRNA 3′ isoform half-lives on a global scale. Our approach combines Anchor-Away of the catalytic subunit of Pol II with direct single-molecule RNA sequencing (DRS). Multiple measurements spread out over a 2 ½-h time frame yield a highly detailed picture of genome-wide 3′ isoform half-lives in yeast. Subsequent bioinformatic analyses can then be used to identify stabilizing and destabilizing elements, potential protein-binding motifs and other interesting features on a global scale.

2 Materials

It is critical to maintain RNase-free conditions for any steps involving RNA handling.

2.1 Growth Media

1. YPD medium: Dissolve 10 g yeast extract, 20 g peptone, and 20 g dextrose in 1 L of H_2O. Autoclave for 20 min.
2. YPD plates: Dissolve 10 g yeast extract, 20 g peptone, and 20 g dextrose in 1 L of H_2O. Add 20 g bacto-agar and autoclave for 20 min. Pour onto 10 cm diameter plates once media has cooled to ~60 °C.

2.2 Reagents

1. 1,000× rapamycin: 1 mg/ml in ethanol. Store at –20 °C for up to 6 months.
2. Acidic phenol-chloroform, pH 4.5.
3. RNA extraction buffer: 10 mM Tris pH 7.5, 10 mM EDTA, 0.5 % SDS.
4. 3 M Sodium acetate.
5. 100 % Ethanol.
6. 70 % Ethanol.
7. RNase-free H_2O.
8. RNase-free DNase Set (recommended vendor QIAGEN).
9. RNeasy Mini kit (recommended vendor QIAGEN).

2.3 Equipment

1. Spectrophotometer.
2. Nanodrop instrument (recommended vendor Nanodrop).

3 Method

3.1 Cell Growth and Time Course

1. Streak out parental strain from a –70 °C stock onto a fresh YPD plate and incubate at 30 °C for 3 days (*see* **Note 1**).
2. Pick a fresh colony and disperse into a 250 ml Erlenmeyer flask containing 50 ml of YPD pre-warmed to 30 °C. Grow overnight at 30 °C on a platform shaker or incubator with vigorous aeration (275 rpm).
3. The next day, seed an appropriate number of cells (*see* **Note 2**) into a 1 l Erlenmeyer flask containing 200 ml of pre-warmed YPD (*see* **Note 3**). Incubate in a 30 °C shaker with vigorous aeration (275 rpm) overnight.
4. The following morning, monitor cell growth by measuring the OD_{600} of the culture on an hourly basis (*see* **Note 2**) until cells reach $OD_{600} = 0.4–0.5$.
5. Once cell culture has reached the desired optical density, harvest 25 ml of the culture by transferring into a 50 ml conical tube. (Optional: add an aliquot of *S. pombe* spike-in control to the harvested cells—*see* **Note 4**.) This is the zero-minute time point.
6. Immediately add 175 μl of 1,000× rapamycin to the growing culture.
7. Spin the harvested cells down at 3,000 rpm (1,800 × *g*) in a table-top centrifuge at room temperature and discard supernatant.
8. Resuspend cells in 1 ml of H_2O and transfer into a standard 1.5 ml microcentrifuge tube that is suitable for phenol–chloroform work (*see* **Note 5**).
9. Spin cells at maximum speed for 5 s.

10. Carefully aspirate away supernatant and immediately freeze the tube containing the pellet in liquid N_2.
11. Precisely 10 min post-rapamycin addition, transfer another 25 ml of cells into a 50 ml conical bottom tube. (Optional: add an aliquot of *S. pombe* spike in to the harvested cells.) This is the 10-min time point.
12. Pellet, wash, and freeze the harvested cells as in **steps 8** to **11**.
13. Repeat the harvest, wash, and freezing steps at the following times (after rapamycin addition): 20 min, 40 min, 1 h, 1 h 20 min, 1 h 40 min, 2 h, and 2½ h.

3.2 Hot Phenol Extraction of RNA

1. Resuspend frozen cell pellets in 400 µl RNA extraction buffer.
2. Add 400 µl acid phenol-chloroform, and mix thoroughly by vortexing for 5–10 s. Incubate for 1 h at 65 °C, with occasional vortexing (every 10–15 min).
3. Place on ice for 5 min.
4. Spin 5 min at top speed in a microfuge at room temperature.
5. Transfer supernatant to a new 1.5 ml microfuge tube suitable for phenol–chloroform work.
6. Add 400 µl acid phenol-chloroform and vortex sample for 5 s.
7. Spin 5 min at maximum speed in a microcentrifuge at room temperature.
8. Transfer 250 µl supernatant into a new 1.5 ml tube.
9. Add 25 µl of 3 M NaOAc and 625 µl of EtOH, vortex briefly, and place at –70 °C for 15 min.
10. Spin for 15 min at top speed in a microfuge at room temperature.
11. Carefully pipet away aqueous phase, leaving approximately 50 µl and the pellet behind.
12. Wash pellet by adding 1 ml of 70 % EtOH. Mix by inversion and spin for 5 min at maximum speed at room temperature.
13. Pipet away and discard as much of the supernatant as possible.
14. Briefly air-dry pellet (<10 min).
15. Resuspend in 300 µl RNase-free H_2O.
16. Determine the concentration with the Nanodrop instrument. Concentration is typically ≥2 µg/µl.

3.3 Removal of Contaminating DNA and Further Purification of Total RNA

1. Into a new microcentrifuge tube, transfer 100 µg RNA, 10 µl buffer RDD (QIAGEN), H_2O up to 100 µl, and 2.5 µl DNase I (QIAGEN). Mix contents by gently pipetting up and down several times but *do not vortex*!
2. Incubate at room temperature for 20 min.

3. Add 350 µl buffer RLT and 250 µl ethanol. Mix by pipetting and load sample onto a QIAGEN RNeasy mini-column.
4. Spin 30 s. at 10,000 rpm (9,400 × *g*). Discard flow-through.
5. Wash with 500 µl buffer RPE. Spin for 30 s at 10,000 rpm and discard flow-through.
6. Repeat wash with 500 µl buffer RPE. Spin for 2 min at 10,000 rpm.
7. Place column into new tube and spin for 1 min at 10,000 rpm.
8. Place column into collection tube and add 30 µl of H_2O directly to the center of the membrane. Let sit at room temperature for 1 min and then spin for 1 min at 10,000 rpm.
9. Pipet an additional 30 µl of H_2O onto the center of the column, let sit at room temperature for 1 min, and repeat 1-min spin at 10,000 rpm.
10. Discard column and measure the concentration of RNA in the eluate by Nanodrop. Keep RNA frozen at −20 or −70 °C.
11. Proceed to single-molecule direct RNA sequencing (*see* **Note 6**).

4 Notes

1. We typically perform half-life measurements in a strain (JGY2000) that carries a tagged version of Rpb1, the catalytic subunit of Pol II, in a BY4741 background [1]. In theory, other genetic backgrounds and Pol II subunits could be used in place of the BY4741/ Rpb1 combination. However, it is imperative to test the efficacy of the Anchor-Away assay in these strains by using chromatin immunoprecipitation (ChIP) to confirm the depletion of various Pol II-transcribed genes after rapamycin addition [12, 14].
2. JGY2000 has a doubling time of ~90 min. We typically seed the equivalent of 2,500 cells/ml (1.2×10^{-4} OD_{600} units/ml; this is usually just a few µl of the overnight culture dispensed into the 200 ml of fresh YPD in **step 3**) in the early evening. The following morning, 16 h later, the cells will have reached $OD_{600} = 0.2$. We then monitor culture growth by taking hourly OD_{600} readings to ensure that the cells are growing rapidly until they are ready for the time course ($OD_{600} = 0.4$–0.5). Initial seeding amounts will almost certainly have to be adjusted if the strain is grown in media other than YPD and if strains of different genetic backgrounds are to be used.
3. Other growth media (e.g., synthetic complete medium, minimal medium) can be used in place of YPD in order to obtain growth condition-specific 3′ isoform half-lives. However, it is essential to verify that the Anchor-Away of Pol II is working

well by measuring Pol II occupancy at a handful of genes that are expressed under the specific growth conditions used.

4. *S. pombe* spike-in controls are very useful for normalizing cross-time point variation due to sample manipulation errors. We typically add a fixed number of *S. pombe* cells (a 100 µl aliquot containing the equivalent of 0.1 OD_{600} of cells) to each time point containing 25 ml of *S. cerevisiae* culture. This results in an approximate 100:1 *S. cerevisiae*:*S. pombe* cell ratio at the zero time point. This ratio will naturally change as the *S. cerevisiae* mRNA is degraded, while the *S. pombe* mRNA level remains constant. Once the overall number of *S. pombe* read counts at each time point is determined, an appropriate correction factor can be applied to the relevant *S. cerevisiae* reads, thereby normalizing read counts at different time points. Other spike-in controls can be used in place of (or in addition to) *S. pombe*. The suitability and quantities of other controls has to be empirically determined on a case-by-case basis.
5. To avoid spills or leaks, take care to use high quality tubes for phenol extraction. Phenol is extremely toxic—wear appropriate lab gear and gloves, and perform extractions in a chemical fume hood.
6. 3′ isoform abundance at different time points is best measured by direct RNA sequencing (DRS) on a Helicos platform as described [15]. In the USA, commercial DRS services are available through SeqLL (www.seqll.com). Alternatives to DRS include 3P-seq [16], TIF-Seq [17], and 3′ READS [18].

References

1. Moqtaderi Z, Geisberg JV, Jin Y et al (2013) Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. Proc Natl Acad Sci U S A 110:11073–11078
2. Sherstnev A, Duc C, Cole C et al (2012) Direct sequencing of Arabidopsis thaliana RNA reveals patterns of cleavage and polyadenylation. Nat Struct Mol Biol 19:845–852
3. Pelechano V, Wei W, Steinmetz LM (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. Nature 497:127–131
4. Schlackow M, Marguerat S, Proudfoot NJ et al (2013) Genome-wide analysis of poly(A) site selection in Schizosaccharomyces pombe. RNA 19:1617–1631
5. Spies N, Burge CB, Bartel DP (2013) 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Res 23:2078–2090
6. Hoque M, Ji Z, Zheng D et al (2013) Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nat Methods 10:133–139
7. Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138:673–684
8. Ulitsky I, Shkumatava A, Jan CH et al (2012) Extensive alternative polyadenylation during zebrafish development. Genome Res 22:2054–2066
9. Tian B, Manley JL (2013) Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci 38:312–320
10. Batra R, Charizanis K, Manchanda M et al (2014) Loss of MBNL leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. Mol Cell 56:311–322

11. Jenal M, Elkon R, Loayza-Puch F et al (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149:538–553
12. Haruki H, Nishikawa J, Laemmli UK (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol Cell 31:925–932
13. Fan X, Geisberg JV, Wong KH et al (2011) Conditional depletion of nuclear proteins by the Anchor Away system. Curr Protoc Mol Biol. doi:10.1002/0471142727.mb1310bs93
14. Geisberg JV, Moqtaderi Z, Fan X et al (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156:812–824
15. Ozsolak F (2014) Quantitative polyadenylation site mapping with single-molecule direct RNA sequencing. Methods Mol Biol. doi: 10.1007/978-1-62703-971-0_13
16. Jan CH, Friedman RC, Ruby JG et al (2011) Formation, regulation and evolution of Caenorhabditis elegans 3′ UTRs. Nature 469:97–101
17. Pelechano V, Wei W, Jakob P et al (2014) Genome-wide identification of transcript start and end sites by transcript isoform sequencing. Nat Protoc 9:1740–1759
18. Jin Y, Geisberg JV, Moqtaderi Z et al (2015) Mapping 3′ mRNA isoforms on a genomic scale. Curr Protoc Mol Biol. doi: :10.1002/0471142727.mb0423s110

Chapter 21

Visualizing mRNA Dynamics in Live Neurons and Brain Tissues

Hye Yoon Park and Minho Song

Abstract

Localization of mRNA plays a crucial role in a variety of neuronal processes including synaptogenesis, axonal guidance, and long-term plasticity. Recent advances in fluorescence imaging and RNA labeling techniques allow us to visualize how individual mRNA molecules are dynamically regulated inside live neurons and brain tissues. Here, we describe key methods in imaging mRNA dynamics, including preparation of neuron culture and brain slices from transgenic mice expressing GFP-labeled mRNA, high-resolution detection of single molecules, live tissue imaging, and analysis of mRNA transport.

Key words RNA localization, Single-molecule imaging, Live cell imaging, MS2-GFP system, Two-photon microscopy

1 Introduction

Local protein synthesis is one of the major mechanisms to regulate gene expression in complex dendritic and axonal processes of neurons with a high spatial and temporal control (for review, *see* Ref. 1). A prerequisite for local translation is the presence of mRNA; thus it has been of great interest to identify mRNA molecules that are transported and targeted to the specific sites for translation. Localization of mRNA is a prevalent phenomenon as illustrated in the CA1 region of the hippocampus, where 2550 out of 8379 transcripts are localized in dendrites and/or axons [2]. While high-resolution fluorescence in situ hybridization (FISH) techniques have revealed the location of individual mRNAs in cultured neurons and brain slices, it is still largely unknown how these mRNA molecules are sorted and delivered to the final destination of local translation.

In order to study RNA dynamics in live cells, various labeling techniques have been introduced including GFP-tagged RNA, oligonucleotide probes, and aptamers [3–5]. Among these, the MS2-GFP system [6] has been widely used to study RNA dynamics

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_21,

in live cells at a single-molecule resolution. The technique employs the high-affinity binding (K_d ~5 nM) of the MS2 bacteriophage capsid protein (MCP) to the sequence-specific RNA stem-loop-binding site (MBS). MCP is fused with green fluorescent protein (MCP-GFP), and the MBS repeats are added into the RNA of interest. As the mRNA is transcribed, an MCP dimer binds to each MS2 stem-loop. Consequently, each target mRNA is labeled with multiple copies of the MCP-GFP allowing single-molecule detection. The benefit of the MS2-GFP system has been demonstrated in several studies of transcription [7–9], RNA transport [10–15], and localization [6, 16–21]. These live-cell studies have provided valuable insights into the behavior of mRNA in real time. However cells in a two-dimensional culture dish have different morphologies than those in the three-dimensional structure, lack normal cell-cell interactions, and hence may not accurately manifest the RNA regulation in the native tissue environment.

In order to visualize mRNA dynamics in live tissues, the MS2-GFP system is recently extended to live mice [22]. A transgenic mouse in which MCP-GFP is ubiquitously expressed under the control of ubiquitin-C promoter (MCP mouse) was generated and crossed with the Actb-MBS mouse [23], a knock-in mouse in which 24 repeats of MBS are inserted into the 3′ untranslated region (UTR) of the β-actin gene. In the hybrid mouse (MCP × MBS mouse), endogenous β-actin mRNAs are labeled with the MS2-GFP system without an adverse effect in live animals. By using double homozygous MCP × MBS mice, the dynamics of all endogenous β-actin mRNA can be monitored in live primary cells and acute brain slices at a single-molecule resolution in real time. We propose that the MS2-GFP mouse system provides a powerful tool to visualize how RNA metabolism is dynamically regulated in the native microenvironment during tissue homeostasis, differentiation, and regeneration. In this chapter, we describe how to visualize endogenous β-actin mRNA in live neurons and brain tissues of MCP × MBS mice and how to analyze the mRNA dynamics from the imaging data.

2 Materials

Use ultrapure water (18 MΩ cm equivalent) in the preparation of all materials.

2.1 Live Neuron Imaging

1. Animals: 0.5–2.5-day-old MCP × MBS mouse pups (*see* **Note 1**).
2. 10× boric acid buffer (BAB): 50 mM boric acid, 12.5 mM sodium tetraborate decahydrate, pH 8.5. Add ~300 mL water to a glass beaker. Weigh 1.24 g boric acid and 1.91 g sodium tetraborate decahydrate and transfer to the beaker. Mix, adjust pH to 8.5, and make up to 400 mL with water.

3. Poly-d-lysine (PDL) solution: 2 mg/mL poly-d-lysine in 10× BAB. Dissolve 20 mg of PDL in 10 mL 10× BAB. Store 1 mL aliquots at −20°C.
4. Neural dissection solution (NDS): 10 mM Hanks' balanced salt solution (HBSS), 10 mM HEPES in water. Add 44.5 mL water in a sterile 50 mL conical centrifuge tube. Add 5 mL 10× HBSS, no calcium, no magnesium, no phenol red, and 0.5 mL 1 M HEPES. Make fresh and keep on ice.
5. Plating medium (PM): 10 % Fetal bovine serum (FBS), 1× Glutamax, 0.1 mg/mL Primocin in Neurobasal A medium. Add 17.8 mL Neurobasal-A medium, 2 mL FBS heat inactivated, 200 µL Glutamax, and 40 µL Primocin in a sterile 50 mL conical centrifuge tube. Make fresh and keep at 37°C for about 30 min before use.
6. B27 medium: 1× B27, 1× Glutamax, 0.1 mg/mL Primocin in Neurobasal A medium. Add 38.7 mL Neurobasal-A medium, 800 µL B27 serum-free supplement, 400 µL Glutamax, and 80 µL Primocin in a sterile 50 mL conical centrifuge tube. Store at 4°C.
7. Trypsin: 2.5 % trypsin without phenol red. Aliquot 200 µL and store at −20°C.
8. HEPES-buffered saline (HBS): 119 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 30 mM d-glucose, 20 mM HEPES at pH 7.4. Add ~150 mL water to a glass beaker. Add 4.76 mL 5 M NaCl, 1 mL 1 M KCl, 0.4 mL 1 M $CaCl_2$, 0.4 mL 1 M $MgCl_2$, 1.08 g d-glucose, and 4 mL 1 M HEPES to the beaker. Mix, adjust pH to 7.4 with ~150 µL 5 N NaOH, and make up to 200 mL with water. Make fresh and keep at 37°C for about 30 min before use.
9. Equipment for dissection: Laminar flow hood, tissue culture incubator (37°C, 5% CO_2 in humidified air), dissection stereoscope, two fine-tipped forceps (Dumont No. 5), small surgical scissors, #10 sterile scalpels, plastic transfer pipets, hemocytometer.
10. 35 mm glass-bottom petri dishes.
11. Sterile plasticware: 5, 10, and 25 mL serological pipets, 50 and 100 mm petri dishes.
12. Equipment for live neuron imaging: Inverted microscope (Olympus IX71 or similar), electron multiplying charge-coupled device (EMCCD) camera (Andor iXon or similar), 150×, 1.45 NA oil immersion objective, live-cell imaging incubator.
13. Image acquisition software: Micro-Manager (one can download Micro-Manager from http://www.micro-manager.org) or commercial software integrated in the microscope system.
14. Image processing software: Fiji (one can download Fiji from http://fiji.sc/Fiji).

2.2 Acute Brain Slice Imaging

1. Animals: 14–21-day-old MCP×MBS mouse (*see* **Notes 1** and **2**).
2. Artificial cerebrospinal fluid (ACSF) solution: 119 mM NaCl, 2.5 mM KCl, 1.3 mM $MgSO_4$, 2.5 mM $CaCl_2$, 1.0 mM NaH_2PO_4, 26.4 mM $NaHCO_3$, 11.0 mM d-glucose. Add 2.5 mL of 1 M $CaCl_2$ in ~800 mL water in a glass beaker and mix well. Add 6.954 g NaCl (MW 58.44), 0.186 g KCl (MW 74.55), 0.157 g $MgSO_4$ (MW 120.37), 0.156 g $NaH_2PO_4°2H_2O$ (MW 155.99), 2.218 g $NaHCO_3$ (MW 84.01), and 1.982 g d-glucose to the beaker. Mix and make up to 1 L with water. Check if osmolarity is 290–310 mOsm. Make fresh and keep on ice for about 20 min before use.
3. Anesthetic: Isoflurane.
4. Fume hood.
5. Bell jar.
6. Vibratome (Leica VT1000S or similar).
7. Carbogen gas (95% O_2/5% CO_2).
8. Surgery tools: Large scissors, small surgical scissors, rongeur, mini hippocampal tool, spatulas, blunt tweezers, filter papers, plastic spoon, small paintbrush, #10 sterile scalpels.
9. Vetbond.
10. Sterile plasticware: Plastic transfer pipets, 50 and 100 mm plastic petri dishes.
11. Prechamber for holding slices.
12. Two-photon laser scanning microscope (2PLSM) with water immersion objective (Olympus XLPLN25XWMP2 25×/1.05 NA or similar).
13. Imaging chamber equipment: Recording chamber, in-line solution heater, heater controller, cable assembly for heater controllers, platform for recording chamber heater, slice hold-down harp, peristaltic pump.
14. Image acquisition software: ScanImage custom written in MATLAB [24] (One can download ScanImage from http://scanimage.org) or commercial software integrated in the microscope system.
15. Image processing software: Fiji (One can download Fiji from http://fiji.sc/Fiji).

3 Methods

3.1 Imaging mRNA in Live Neurons

3.1.1 Poly-d-Lysine (PDL) Coating of Glass-Bottom Dishes

1. Thaw a vial of PDL stock, make a 1:10 dilution (0.2 mg/ml) in water, and filter sterilize it using a syringe filter (0.45 μm pore) into a sterile 15 mL conical centrifuge tube.
2. Add 200 μL of 0.2 mg/mL PDL solution onto the glass surface of each glass-bottom petri dish and incubate in the tissue culture incubator (37°C, humidified) overnight.

3. Remove PDL solution, wash the glass-bottom dishes with sterile water, remove water, and keep them in the tissue culture incubator before use.

3.1.2 Dissection of Hippocampi

1. Sacrifice 3–4 mouse pups (*see* **Note 3**) using a method approved by the institutional animal care and use committee (*see* **Note 1**).
2. In the laminar flow hood, remove the brains and place each brain on a filter paper in ice-cold neural dissection solution (NDS) in a 50 mm petri dish. Keep them on ice for 5 min.
3. Under a dissection stereoscope, remove the hindbrain with a sterile scalpel and carefully peel off the meninges with fine-tipped forceps (*see* **Note 4**).
4. Split the hemispheres using a scalpel and dissect the hippocampi using forceps.
5. Using a wide-bore plastic transfer pipet, transfer dissected hippocampi into ice-cold NDS in a 15 mL conical centrifuge tube and let them settle down to the bottom of the tube on ice.

3.1.3 Dissociation and Plating

1. Bring the tube containing hippocampi to the laminar flow hood and carefully aspirate NDS until ~2 mL of NDS is left in the tube.
2. Add 200 μL of 10× trypsin and incubate the tube in 37°C water bath for 15 min.
3. Remove trypsin, add 3 mL plating medium (PM), and triturate with 5 mL pipet 20–30 times until most large chunks are dissociated. Be careful not to generate too many bubbles.
4. Wait for ~3 min until undigested tissues settle down to the bottom of the tube.
5. Determine the cell density in the suspension using a hemocytometer.
6. Resuspend the cells to 425,000 cells/mL in PM.
7. Seed 200 μL of resuspended cells on the PDL-coated glass part of each glass-bottom dish.
8. Plate cells in the incubator for 4 h.
9. Equilibrate B27 medium in a sterile petri dish in the incubator for 2–4 h.
10. Gently add 1.8 mL of equilibrated B27 medium to each glass-bottom dish.

3.1.4 Culture Maintenance

1. Equilibrate B27 medium in the incubator for a few hours or overnight before feeding the cells.
2. At 4, 7, 14, and 21 days after seeding, add 300 μL of B27 medium to each dish.

3.1.5 Imaging mRNA in Live Neurons

1. Turn on the temperature controller and the humidifier for the imaging chamber and wait for ~30 min to equilibrate the chamber at 37°C.
2. Remove B27 medium from neurons grown in a glass-bottom dish and carefully wash the neurons twice with pre-warmed (37°C) HEPES-buffered saline (HBS). Add 2 mL of fresh HBS in the dish and mount onto the inverted microscope stage.
3. Using a high-magnification objective lens (150×), look for dendrites that contain mRNA particles.
4. Start time-lapse imaging with an interval and duration appropriate for the mRNA dynamics being analyzed. For mRNA transport in neurons, we typically take images every 100 ms with streaming for 1 min (600 frames in total).

3.2 Imaging mRNA in Acute Brain Slices

3.2.1 Preparation of Acute Brain Slices

1. Prepare artificial cerebrospinal fluid (ACSF) and keep it on ice.
2. Oxygenate ACSF in a beaker for ~30 min on ice using a bubbler and tubing connected to the carbogen gas (95% O_2/5% CO_2) tank.
3. Euthanize a mouse using a method approved by the institutional animal care and use committee (*see* **Note 1**). One method is to anesthetize a mouse with isoflurane in a bell jar and decapitate the mouse with sharp scissors (*see* **Note 5**).
4. Immediately after decapitation, place the head on folded paper towels. Using a #10 scalpel, make an incision along the midline of the scalp.
5. Flip the skin and cut the skull along the midline from the neck to between the eyes using small surgical scissors.
6. Using a rongeur, peel the skull on both sides.
7. Quickly remove the brain and place it in the beaker with oxygenated ice-cold ACSF for ~5 min.
8. Turn on the vibratome and fill the reservoir outside the cutting chamber with ice.
9. Fill a 100 mm petri dish with ice, place a filter paper on the cover of the petri dish, and wet the filter paper completely with oxygenated ice-cold ACSF.
10. Retrieve the brain from the beaker with a plastic spoon and place it on the wet filter paper.
11. Using a scalpel, remove the hindbrain and approximately one-quarter of the frontal lobes. Cut the brain into two hemispheres.
12. Apply a drop of vetbond on the cutting stage, and glue each hemisphere with front side down.

13. Fill the cutting chamber with oxygenated ice-cold ACSF until two hemispheres are fully immersed.
14. Mount a blade and bring the cutting chamber up until the blade is immersed in ACSF. Cut 350–500 μm thick coronal sections at slow speed (~0.2 mm/s) and high vibration frequency (~80 Hz).
15. Collect slices with a wide-bore plastic transfer pipet or a small paint brush and transfer them into a 50 mm petri dish containing oxygenated ice-cold ACSF.
16. Transfer slices to the prechamber and recover them at room temperature in oxygenated ACSF for about 2 h.

3.2.2 Two-Photon Imaging of Brain Slices

1. On the two-photon laser-scanning microscope, prepare the recording chamber maintained at 32°C and continuously perfused with warm oxygenated ACSF at a flow rate of ~1 mL/min.
2. Transfer a brain slice to the recording chamber and secure the position of the slice by using a slice hold-down harp.
3. Using a water immersion low-magnification objective lens with 880 nm excitation wavelength, look for the brain regions of interest (e.g., neocortex, hippocampus).
4. Choose the regions of interest (ROIs) from the low magnification overview and zoom in to satisfy the Nyquist sampling criteria (*see* **Note 6**). Take time-lapse *z*-stack images with an interval and duration appropriate for the mRNA dynamics being analyzed.

3.3 Analyzing mRNA Transport

3.3.1 Selecting the Region of Interest and Straightening

1. Open the time-stack image file using Fiji (in the main menu "File" > "Open").
2. To select a region of interest (e.g., along the dendrite), click the line tool button on the toolbar. Left click twice on the line tool button and set the line width to approximately 30–50 pixels (*see* **Note 7**). Right click on the line tool button and select "Segmented Line."
3. Draw a segmented line along the region of interest (*see* Fig. 1a). When you left click once, a new line is started. Create a segmented line selection by left clicking at connecting points. Right click or double click to finish the line. To straighten the segmented line, select the "Edit" > "Selection" > "Straighten…" menu option. In the pop-up window, enter a file name and line width, check "Process Entire Stack" option, and click "OK."
4. Now, you can see the straightened image of the region you selected (*see* Fig. 1b and **Note 8**). The starting point of the line is on the left side of the straightened image. Save the image (*see* **Note 9**).

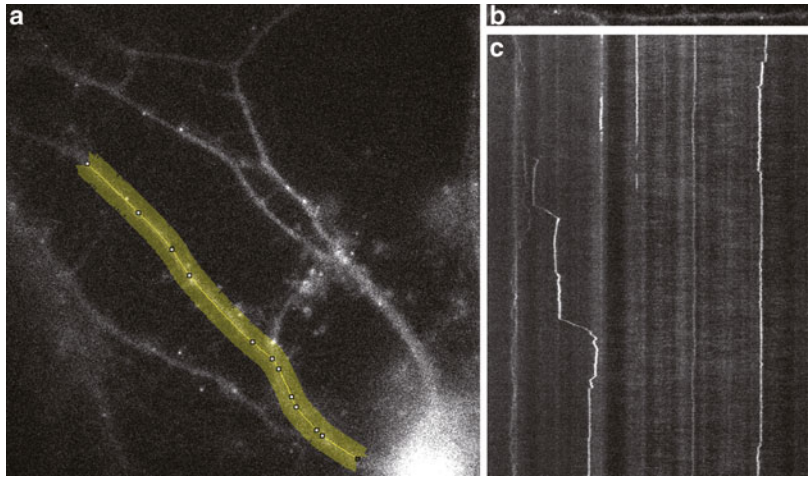

Fig. 1 Image analysis of mRNA transport in live neurons. (**a**) A representative image of a neuron cultured from MCP × MBS mouse at 16 days in vitro. The *yellow-shaded* area indicates a line selection along the dendrite of interest. (**b**) A straightened image of the dendrite selected in (**a**). (**c**) A kymograph generated from the straightened time-stack image shown in (**b**)

3.3.2 Generating Kymographs

1. Analysis of particle transport can be facilitated by generating kymographs. First, generate a resliced image stack by selecting "Image" > "Stacks" > "Reslice [/]…." In the pop-up window, choose "start at Top," uncheck "Flip vertically" and "Rotate 90°," check "Avoid interpolation." and click "OK."
2. Click the resliced image stack and select "Image" > "Stacks" > "Z Project…." In the pop up window, enter "1" in the "Start slice" and the last slice number in the "Stop slice." Choose the option in the "Projection type" ("Max Intensity" or "Standard Deviation" will be the best choice). Then, Press "OK." Now, you can see the kymograph of the straightened time-stack image. The *x*-axis and -*y*-axis correspond to the length and time, respectively. Save this kymograph (*see* Fig. 1c).
3. If it is difficult to distinguish the particle trajectories, adjust the contrast by selecting "Image" > "Adjust" > "Brightness/Contrast…" in the main menu.
4. You can easily calculate the velocity and travel distance of the particles by using the kymograph. To distinguish mRNA particles from other background signals, compare the straightened image and the kymograph.

4 Notes

1. All experiments using animals should be carried out under the approval of the institutional animal care and use committee (IACUC).

2. It is desirable to use younger mice because autofluorescence from lipofuscin in lysosomes [25, 26] complicates single mRNA imaging in adult mouse tissues.
3. Hippocampi from 3 to 4 pups will typically yield ~16 glass-bottom dishes at 85,000 neurons per dish.
4. Remove the meninges thoroughly during dissection since fibroblasts in the meninges can overgrow neurons in mixed cultures. More details on hippocampal dissection can be found in Refs. [27, 28].
5. Different protocols of preparing acute brain slices can be found in reference [29–31]
6. For imaging individual mRNP particles, we typically take images with an Olympus XLPLN25XWMP2 25×/1.05 NA water immersion objective. The 1/e radii (ω) of the lateral (xy) and axial (z) intensity-squared profiles [32] at 880 nm excitation are estimated to be 190 nm and 640 nm, respectively. To meet the Nyquist sampling criteria, a pixel size should be ~190 nm and a *z*-step interval should be ~640 nm.
7. The line width must be set up adequately; not too narrow, not too wide. If the line selection is too wide, you could over-detect particles (e.g., the particles in other cells or dendrites).
8. During this process, the image can be distorted. Selecting one straight line is the best way to avoid image distortion.
9. To reduce the file size, select "Image" > "Type" > "16-bit" or "8-bit."

Acknowledgements

This work was supported by Research Resettlement Fund for the new faculty of Seoul National University to H.Y.P.

References

1. Jung H, Gkogkas CG, Sonenberg N, Holt CE (2014) Remote control of gene function by local translation. Cell 157:26–40
2. Cajigas IJ, Tushev G, Will TJ, Dieck ST, Fuerst N, Schuman EM (2012) The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron 74:453–466
3. Tyagi S (2009) Imaging intracellular RNA distribution and dynamics in living cells. Nat Methods 6:331–338
4. Rodriguez AJ, Condeelis J, Singer RH, Dictenberg JB (2007) Imaging mRNA movement from transcription sites to translation sites. Semin Cell Dev Biol 18:202–208
5. Paige JS, Wu KY, Jaffrey SR (2011) RNA mimics of green fluorescent protein. Science 333:642–646
6. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2:437–445
7. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123: 1025–1036

8. Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) Transcriptional pulsing of a developmental gene. Curr Biol 16:1018–1025
9. Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH (2007) In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 14:796–806
10. Rook MS, Lu M, Kosik KS (2000) CaMKII alpha 3′ untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage. J Neurosci 20:6385–6393
11. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH, Bertrand E (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13:161–167
12. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, Spector DL, Singer RH (2004) Dynamics of single mRNPs in nuclei of living cells. Science 304:1797–1800
13. Dynes JL, Steward O (2012) Arc mRNA docks precisely at the base of individual dendritic spines indicating the existence of a specialized microdomain for synapse-specific mRNA translation. J Comp Neurol 520:3105–3119
14. Jaramillo AM, Weil TT, Goodhouse J, Gavis ER, Schupbach T (2008) The dynamics of fluorescently labeled endogenous gurken mRNA in Drosophila. J Cell Sci 121:887–894
15. Grunwald D, Singer RH (2010) In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. Nature 467:604–607
16. Beach DL, Salmon ED, Bloom K (1999) Localization and anchoring of mRNA in budding yeast. Curr Biol 9:569–578
17. Forrest KM, Gavis ER (2003) Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol 13:1159–1168
18. Weil TT, Forrest KM, Gavis ER (2006) Localization of bicoid mRNA in late oocytes is maintained by continual active transport. Dev Cell 11:251–262
19. Haim L, Zipor G, Aronov S, Gerst JE (2007) A genomic integration method to visualize localization of endogenous mRNAs in living yeast. Nat Methods 4:409–412
20. Mili S, Moissoglu K, Macara IG (2008) Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. Nature 453:115–119
21. Zimyanin VL, Belaya K, Pecreaux J, Gilchrist MJ, Clark A, Davis I, St Johnston D (2008) In vivo imaging of oskar mRNA transport reveals the mechanism of posterior localization. Cell 134:843–853
22. Park HY, Lim H, Yoon YJ, Follenzi A, Nwokafor C, Lopez-Jones M, Meng XH, Singer RH (2014) Visualization of dynamics of single endogenous mRNA labeled in live mouse. Science 343:422–424
23. Lionnet T, Czaplinski K, Darzacq X, Shav-Tal Y, Wells AL, Chao JA, Park HY, de Turris V, Lopez-Jones M, Singer RH (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods 8:165–U196
24. Pologruto TA, Sabatini BL, Svoboda K (2003) ScanImage: flexible software for operating laser scanning microscopes. Biomed Eng Online 2:13
25. Jung T, Hohn A, Grune T (2010) Lipofuscin: detection and quantification by microscopic techniques. Methods Mol Biol 594:173–193
26. Schnell SA, Staines WA, Wessendorf MW (1999) Reduction of lipofuscin-like autofluorescence in fluorescently labeled tissue. J Histochem Cytochem 47:719–730
27. Kaech S, Banker G (2006) Culturing hippocampal neurons. Nat Protoc 1:2406–2415
28. Neumann S, Campbell GE, Szpankowski L, Goldstein LS, Encalada SE (2014) Characterizing the composition of molecular motors on moving axonal cargo using "cargo mapping" analysis. J Vis Exp 30(92):e52029
29. Debanne D, Boudkkazi S, Campanac E, Cudmore RH, Giraud P, Fronzaroli-Molinieres L, Carlier E, Caillard O (2008) Paired-recordings from synaptically coupled cortical and hippocampal neurons in acute and cultured brain slices. Nat Protoc 3:1559–1568
30. Mathis DM, Furman JL, Norris CM (2011) Preparation of acute hippocampal slices from rats and transgenic mice for the study of synaptic alterations during aging and amyloid pathology. J Vis Exp 23(49): pii: 2330
31. Lein PJ, Barnhart CD, Pessah IN (2011) Acute hippocampal slice preparation and hippocampal slice cultures. Methods Mol Biol 758:115–134
32. Zipfel WR, Williams RM, Webb WW (2003) Nonlinear magic: multiphoton microscopy in the biosciences. Nat Biotechnol 21:1369–1377

Chapter 22

Single-Molecule Live-Cell Visualization of Pre-mRNA Splicing

Robert M. Martin, José Rino, Ana C. de Jesus, and Maria Carmo-Fonseca

Abstract

Microscopy protocols that allow live-cell imaging of molecules and subcellular components tagged with fluorescent conjugates are indispensable in modern biological research. A breakthrough was recently introduced by the development of genetically encoded fluorescent tags that combined with fluorescence-based microscopic approaches of increasingly higher spatial and temporal resolution made it possible to detect single protein and nucleic acid molecules inside living cells. Here, we describe an approach to visualize single nascent pre-mRNA molecules and to measure in real time the dynamics of intron synthesis and excision.

Key words Single molecule, Live-cell imaging, Spinning disk confocal microscopy, Splicing, Pre-mRNA

1 Introduction

The constant development of instruments and techniques for microscopic imaging has been a hallmark of biological research. Starting more than 300 years ago, when Anton van Leeuwenhoek observed for the first time unicellular organisms using single-lens microscopes that he himself designed and constructed, optical microscopes have since been steadily evolving. During the nineteenth century compound microscopes allowed biologists to make the seminal discoveries that culminated in the cell theory [1–3]. By the end of the 1980s, with the development of lasers and the widespread use of fluorescent biological markers, the confocal microscope patented by Marvin Minsky back in 1953 become one of the most exciting instruments available to cell biologists [4]. During the 1990s, the discovery and development of the green fluorescent protein (GFP) and related fluorescent proteins made it possible to genetically tag almost any protein with minimal invasiveness [5–7]. Following this advent, a method was developed to image chromatin dynamics in live cells. Bacterial lac operator repeats were

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8_22,

introduced into the genome of yeast and mammalian cells that expressed a GFP-lac repressor fusion protein. Upon binding of the fluorescent repressor to its target sequence, that particular region of chromatin becomes fluorescent and thus visible in the nucleus [8]. A similar approach was later introduced to visualize RNAs in living cells by genetically inserting the binding sites for the MS2 bacteriophage coat protein in the RNA of interest [9]. The resulting reporter gene was integrated in the genome of cells that express the MS2 coat protein fused to a fluorescent protein. Insertion of the MS2-binding sites in the terminal exon of reporter genes revealed kinetic properties of the entire mRNA life cycle, from transcription to transport in the nucleus and export to the cytoplasm [10, 11], whereas intronic insertions of MS2-binding sites have been used to track the dynamics of splicing [12–14].

Here, we describe a strategy to visualize in real time synthesis and excision of intronic sequences from single pre-mRNA molecules in live cells [12]. The protocol starts with construction of a reporter gene tagged with an array of 24 copies of the MS2-binding site inserted in either an intron or terminal exon. The reporter gene is integrated as a single copy in the genome of human cells by site specific homologous recombination [15, 16]. Cells are transiently transfected to express MS2 coat protein fused to a fluorescent protein. As soon as MS2-binding sites are transcribed, they rapidly bind fluorescent coat proteins [17], making the nascent pre-mRNA molecules visible as diffraction limited objects (*see* **Note 1**). Cells are imaged alive in a spinning disk confocal microscope equipped with a single-photon-sensitive electron-multiplying charge-coupled device (EMCCD) camera.

2 Materials

Prepare all solutions at room temperature unless otherwise noted. Solutions used to culture and prepare cells should be warmed to 37 °C prior to use. All solutions and materials in contact with live cells must be sterile and have to be handled inside a tissue culture hood. Disposal of reagents, solutions and cell material has to be carried out following the specific regulations in effect.

2.1 Cells and Materials for Cell Culture

1. Flp-In™ T-REx™-293 cell line (Thermo Fisher Scientific, Invitrogen).
2. High-glucose (4.5 g/l) Dulbecco's modified essential medium (DMEM), 500 ml. Store at 4 °C.
3. Phenol red free DMEM/F-12 (1:1), F-12 Nutrient mixture (Ham) + glutamine + 15 mM HEPES, 500 ml. Store at 4 °C.
4. Fetal bovine serum (FBS): Aliquots of 50 ml are stored at −20 °C.

5. Standard medium (DMEM+10 % FBS): DMEM (**item 2**) supplemented with 10 % FBS (**item 4**). Prepare a 500 ml bottle under sterile conditions and store at 4 °C.
6. Imaging medium (DMEM/F-12 + 10 % FBS): DMEM/F-12 (**item 3**) supplemented with 10 % FBS (**item 4**). Prepare a 500 ml bottle under sterile conditions and store at 4 °C.
7. Phosphate-buffered saline (PBS): 1×, prepared from 10× sterile stock solution.
8. Trypsin solution.
9. 0.01 % Poly-l-lysine solution.
10. Doxycycline: Dissolved in water, 1 mg/ml stock solution prepared fresh from doxycycline powder.
11. Round glass cover slips: 0.17 mm thick, 30 mm diameter.
12. General cell culture material: cell culture-grade T_{25} and T_{75} flasks with non-filter screw caps and p35 dishes, sterile glass or plastic pipettes, micropipettes, and appropriate sterile tips.

2.2 Materials for Cell Transfection

1. Flp-In™ T-REx™ System plasmid kit (Thermo Fisher Scientific, Invitrogen). The kit includes pcDNA5/FRT/TO vector backbone and plasmid pOG44 encoding the Flp-recombinase.
2. Plasmid containing 24 tandem copies of MS2 stem loops [9] (Addgene plasmid # 31865).
3. Plasmid encoding MS2 coat protein fused to either GFP [18] (Addgene plasmid # 27121) or mCherry [19] (Addgene plasmid # 45930) (*see* **Note 2**).
4. Transfection reagent: X-tremeGene 9 (Roche).
5. Transfection reagent: Lipofectamine® 2000 (Life Technologies).
6. OptiMEM® Reduced Serum Medium (Life Technologies).
7. 10 mg/ml Blasticidin in HEPES 20 mM. Stock stored at −20 °C; working solution stored at 4 °C.
8. 50 mg/ml Hygromycin in PBS. Aliquots stored at 4 °C.

2.3 Microscopy

1. We use the spinning disk confocal imaging system (*see* **Note 3**) commercially provided by 3i (Intelligent Imaging Innovations, Inc.). The 3i Marianas SDC system includes an Axio Observer Z1 inverted microscope (Carl Zeiss) equipped with a Yokogawa CSU-X1 spinning-disk confocal head (Yokogawa Electric) and 100 mW solid-state lasers coupled to an acoustic-optical tunable filter (AOTF). Images are acquired using the 100× (Plan-Apo, 1.4 NA) oil immersion objective under control of Slidebook software (Intelligent Imaging Innovations). Digital images (16-bit) are obtained using a back thinned air-cooled EMCCD camera (Evolve 512, Photometrics). The axial position of the sample is controlled with a piezo-driven stage. The

microscope setup can be seen in Fig. 1. An important feature of this system is blanking of the illuminating laser light in the time between camera exposures through TTL synchronization electronics. This significantly reduces exposure of cells to laser light during image acquisition, thus minimizing photo-damage.

2. The whole microscope body excluding the laser module, camera and spinning disk system is housed in a plexi glass environmental chamber (Pecon). An additional incubation chamber (Pecon) is mounted on the microscope stage and connected to air/CO_2, and humidity controllers. Temperature in both microscope and stage incubation chambers are controlled by a common unit. The environment in both chambers is set at 37 °C with 5 % CO_2 and 100 % humidity.
3. Cells are grown on cover slips that are placed in a dedicated metal chamber (Fig. 2).
4. For image analysis we use the MatLab Compiler Runtime (MCR) version 8.3 (The MathWorks, Inc. http://www.mathworks.com/products/compiler/mcr/). Algorithms were developed [12] that are available upon request.

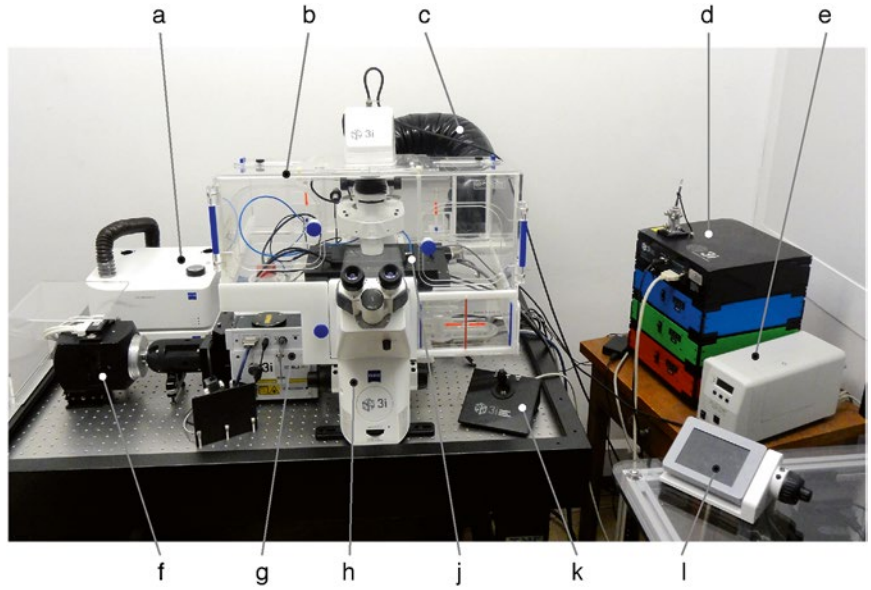

Fig. 1 Spinning disk microscope setup. The indicated components are (**a**) module for stage incubator heating and Air/CO_2 mixture control; (**b**) environmental chamber; (**c**) air inflow for heating the environmental chamber; (**d**) laser module; (**e**) mercury lamp for epifluorescence; (**f**) EMCCD camera; (**g**) spinning disk unit; (**h**) microscope; (**j**) stage incubator; (**k**) *x-y* driving control unit; (**l**) microscope control unit

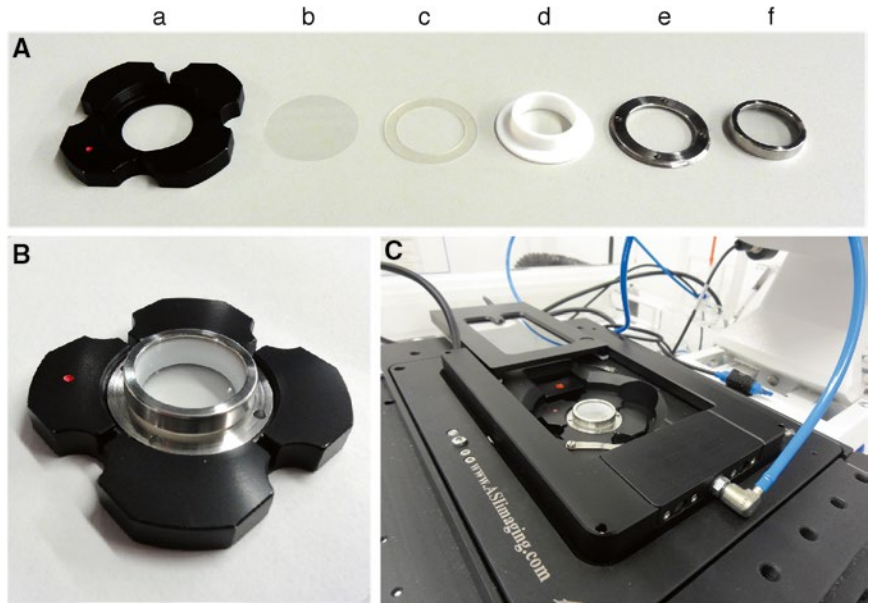

Fig. 2 Cell chamber. (**a**) Disassembled parts of the Pecon POCmini-2 cell chamber: (*a*) base plate; (*b*) 30 mm cover slip; (*c*) silicone ring; (*d*) adapter; (*e*) screwing ring; (*f*) cover. (**b**) Assembled chamber. (**c**) Cell chamber inside the stage incubator. Tubes connect the incubator to the heating, air/CO_2, and humidifying systems

3 Methods

3.1 Cell Culture

3.1.1 Maintaining Cells in Culture

1. Flp-In™T-REx™ -293 cells are grown in T_{25} flasks, until they reach confluence. Cells are kept in culture for no longer than 8–9 passages.
2. Flp-In™T-REx™-293 cells with an integrated transgene are grown in standard medium (**item 5** in Cells and Materials for Cell Culture) supplemented with blasticidin (final concentration 15 μg/ml) and hygromycin (final concentration 200 μg/ml).
3. To split cells, remove medium and rinse with 1× PBS at 37 °C.
4. Apply 0.5 ml of trypsin solution at 37 °C directly on to the surface of the cell layer and tilt gently. Leave flask in the incubator for approximately 1 min.
5. When all cells have detached, add 4.5 ml of medium to the flask and resuspend the cells by gentle pipetting up and down several times. Avoid creating bubbles and foam. Transfer between 0.5 and 1 ml to a new T_{25} flask.
6. Fill up to 5 ml total volume with medium.
7. Place flasks in cell culture incubator and examine the cells daily using an inverted microscope.

3.1.2 Genomic Integration of the Tagged Transgene

1. Insert 24 tandem copies of MS2 stem loops in either an intron or the terminal exon of a reporter gene (*see* **Note 4**). This will be referred to as the tagged transgene.
2. Clone the tagged transgene in the pcDNA5/FRT/TO vector backbone, which contains a Flp-In™ recombination site. This vector also has a eukaryotic expression system with a CMV promoter under control of two copies of the Tet-operator [20].
3. Before transfection, Flp-In™T-REx™ -293 cells are grown in standard medium (**item 5** in Cells and Materials for Cell Culture) supplemented with blasticidin (final concentration 15 μg/ml).
4. In preparation for transfection, cells are seeded at a ratio of 1:6 into two T_{25} flasks and cultured for 2 or 3 days until they reach 50–60 % confluence.
5. In the two T_{25} flasks to be used for transfection, remove the standard medium and add 5 ml of OptiMEM® at 37 °C without antibiotics.
6. Gently vortex the X-tremeGENE 9 reagent (pre-warmed to room temperature) for 30 s just before use.
7. Dilute the pOG44 plasmid encoding the Flp-recombinase to 1 μg/μl (in sterile water).
8. Dilute the pcDNA5/FRT/TO plasmid containing the transgene to 0.1 μg/μl (in sterile water).
9. Prepare two sterile 1.5 ml reaction tubes in the hood.
10. In one tube, add 350 μl OptiMEM®, 9 μl of X-tremeGENE 9, and 2.7 μl of pOG44 plasmid.
11. In the other tube, add 350 μl OptiMEM®, 9 μl of X-tremeGENE 9, 2.7 μl of pOG44 plasmid, and 3 μl of pcDNA5/FRT/TO plasmid.
12. Mix by gentle pipetting and incubate for 15 min at room temperature.
13. Add each transfection solution to one T_{25} flask in a dropwise manner. Shake the flask gently.
14. Place the T_{25} flasks back in the incubator for at least 12 h.
15. Remove the OptiMEM® from the T_{25} flasks and add 5 ml of standard medium without antibiotics.
16. Incubate overnight.
17. Prepare all materials needed to split cells. After trypsin treatment, resuspend the cells in 4.5 ml standard medium and transfer the content of each T_{25} flasks to a T_{75} flask. Fill up to 10 ml total volume with standard medium and add 15 μl blasticidin (from 10 mg/ml stock) and 40 μl hygromycin (from

50 mg/ml stock). This starts the selection process for cell clones with stable integration of the transgene.

18. Place the flasks in the incubator and observe cells daily.
19. After 2 days, aspirate the medium carefully and add 10 ml standard medium with 15 μg/ml blasticidin and 200 μg/ml hygromycin. This procedure has to be repeated every 2–3 days.
20. About 1 week after starting the selection with hygromycin, cells without integration of the transgene will start to die and will be discarded during medium exchange.
21. After about 2 weeks, the control flask transfected only with pOG44 plasmid should be devoid of healthy cells attached to the surface. At this point the control flask can be discarded.
22. Continue to change the medium and antibiotics in the flasks containing cells transfected with the transgene.
23. During the following 1–2 weeks surviving colonies should become visible by eye as small opaque patches when holding the flask in front of a light source. Confirm the colonies with the microscope. Continue to supply the cells with fresh growth medium and antibiotics every 3 days. Use great care in order to avoid losing surviving cells.
24. When the dividing cells in a colony do not spread to the side of a colony but start to grow on top of each other, the cells have to be transferred.
25. Circle the position of the colonies with a lab marker pen on the flask. Aspirate the medium, wash with 7 ml 1xPBS and add 0.5 ml trypsin solution. Tilt flask to distribute the trypsin solution over the complete surface.
26. After 1 min in the incubator, shake flask against the palm of the hand to detach cells. Observe the position of the colonies to make sure all cells have detached.
27. Add 4.5 ml of fresh standard medium and pipet up and down to rinse the entire surface of the flask and make sure all cells are recovered.
28. Transfer the complete volume to a T_{25} flask, fill up to 5 ml and add 15 μg/ml blasticidin and 200 μg/ml hygromycin.
29. Observe cells daily until they reach confluence. Proceed with the maintenance protocol as described in Maintaining Cells in Culture. Prepare frozen cell stocks and store in liquid nitrogen.
30. Test the cell line for expression of the transgene by RT-PCR or fluorescence microscopy. To induce expression of the transgene, add 0.1–1 μg/ml doxycycline to the growth medium and incubate for at least 3 h.

3.2 Preparation of Cells for Imaging Experiments

Timeline of procedures: (1) Split cells on glass cover slips in p35 dishes and incubate overnight. (2) Transfect with MS2 coat protein fused to a fluorescent protein and incubate overnight. (3) Proceed to the microscope on the following day.

3.2.1 Preparation of Cover Slips

1. Place 30 mm cover slips in a dedicated holder inside a closed container and immerse in 70 % ethanol. Place the container in ultrasonic cleaning bath for 30 min.
2. Remove holder with cover slips from ethanol and place in a drying oven. Store the holder with cover slips in a closed container.
3. From this point onwards, all work is carried out inside the cell culture hood. Rinse a forceps with 70 % alcohol and let it dry in the hood. Place p35 plastic dishes (one for each imaging experiment) in the hood.
4. Use the forceps to transfer one cover slip from the holder into one p35 dish.
5. Rinse the dish twice with 1 ml sterile water, making sure that both surfaces of the cover slip are washed.
6. Apply 0.5 ml of poly-l-lysine solution and shake gently, making sure that the entire upper surface of the cover slip is covered. Let it stand in the hood for approximately 1 min. Recover the poly-l-lysine solution for further use.
7. Add 1 ml growth medium (DMEM + 10 % FBS) and proceed to Culturing Cells on Cover Slips.

3.2.2 Culturing Cells on Cover Slips

1. Cells are grown in T_{25} flasks until they reach confluence. Cells are then detached with trypsin as described in Subheading 3.1.1.
2. Once cells have detached (Maintaining Cells in Culture, **step 8**), add 4.5 ml standard medium and resuspend cells as described (Maintaining Cells in Culture, **step 9**).
3. Aspirate all growth medium from the dishes with cover slips and let them stand in the hood for approximately 1 min.
4. Seed 0.5 ml to 1 ml from the cell suspension (*see* **Note 5**) and fill up to 2 ml standard medium. Shake gently to ensure homogeneous distribution of cells.
5. Incubate at 37 °C overnight.

3.2.3 Transfection of Cells

1. In preparation for the transfection procedure: Warm up the frozen plasmid stock, OptiMEM® and Lipofectamine 2000 reagent to room temperature.
2. For each p35 dish to be transfected, prepare and label two sterile 1.5 ml reaction tubes in the hood.
3. Vortex the Lipofectamin 2000 stock solution for 15 s.

4. Mix the plasmid stock solution by tipping against the tube.
5. Pipette 50 µl of OptiMEM into each tube.
6. To one tube, add 3 µl Lipofectamin 2000.
7. To the other tube, add 1 µg DNA plasmid encoding MS2 coat protein (1 µg/µl).
8. Incubate for 5 min at room temperature.
9. Combine the content of the two tubes and mix by pipetting once up and down; incubate for 20 min at room temperature.
10. Add the transfection solution drop wise to the p35 dish with cells grown on a cover slip. Shake gently.
11. Place the p35 dish back in the incubator at 37 °C and incubate overnight.

3.2.4 Induction of Transgene Expression

1. Around 3 h before going to the microscope, warm up microscopy medium (Phenol red free DMEM/F-12 mix + 10 % FBS) to 37 °C.
2. In each dish, remove the standard medium (DMEM + 10 % FBS) and add 2 ml microscopy medium (Phenol red free DMEM/F-12 mix + 10 % FBS).
3. Add 2 µl doxycycline (0.1 mg/ml stock solution; final concentration 0.1 µg/ml).
4. Shake gently and place the dish back in incubator at 37 °C for at least 3 h.

3.2.5 Assembly of Cell Chamber

1. All parts of the cell chamber are washed in deionized water and dried before use (*see* Fig. 2a for details of the components).
2. Using a forceps, transfer the cover slip with cells from the p35 dish the base plate of the chamber (this can be done in a hood but in our experience it is not required).
3. Take the silicone ring with forceps and place it on the cover slip inside the base plate; mount sequentially the adapter and the screwing ring. Tighten the screwing ring carefully to seal the chamber but pay special attention to avoid breaking the cover slip. The assembled chamber is shown in Fig. 2b.
4. Transfer 1 ml of medium from the p35 dish to the chamber. Use a micropipette tip touching the wall of the chamber. Add medium slowly and carefully to avoid damaging cells on the cover slip due to flow forces.
5. Use the cover ring with glass window to close the chamber.
6. Place the chamber as fast as possible inside the environmental chamber of the microscope (Note: The microscope and incubation systems should be switched on at least 2 h before).

3.3 Cell Imaging

1. Start all microscope components and imaging software.
2. Select the 100 × objective.
3. Apply one drop of immersion oil to the objective, without touching the lens.
4. Mount the chamber with cells inside the stage incubator (Fig. 2c).
5. Allow for temperature to equilibrate for 15 min.
6. Use bright-field illumination to focus the cells.
7. Switch off the bright-field illumination and change the light path to the confocal unit.
8. Adjust all settings as described in Table 1.
9. Observe cells in the live camera mode. Avoid exposing cells to unnecessary fluorescence excitation radiation to reduce photo-damage.
10. Search for cells with low levels of nuclear fluorescence and clear diffraction limited objects (or "dots"). For reporter genes containing MS2 binding sites inserted in an intron, a single immobilized dot should be visible in the nucleus (Fig. 3a). We refer to this type of dot as the "transcription dot" because it corresponds to nascent pre-mRNA molecules still associated with the DNA template. In contrast, when MS2-binding sites are inserted in the terminal exon, a multitude of dots are visible diffusing throughout the nucleus (Fig. 3c) and the cytoplasm. Each diffusing dot in the nucleus corresponds to a single-

Table 1

Settings for live-cell imaging

Laser	100 mW solid-state laser
	Intensity is reduced to 30 % of maximum
	Excitation for GFP: 488 nm
EMCCD	Intensification: 30% of maximum
	Speed:1
	Gain: 1
	Binning: 1 × 1
	Image size: 512 × 512 pixel
	(*xy*-pixel size: 0.212 µm)
3D acquisition	*z*-step size: 0.27 µm
Exposure time	30–50 ms

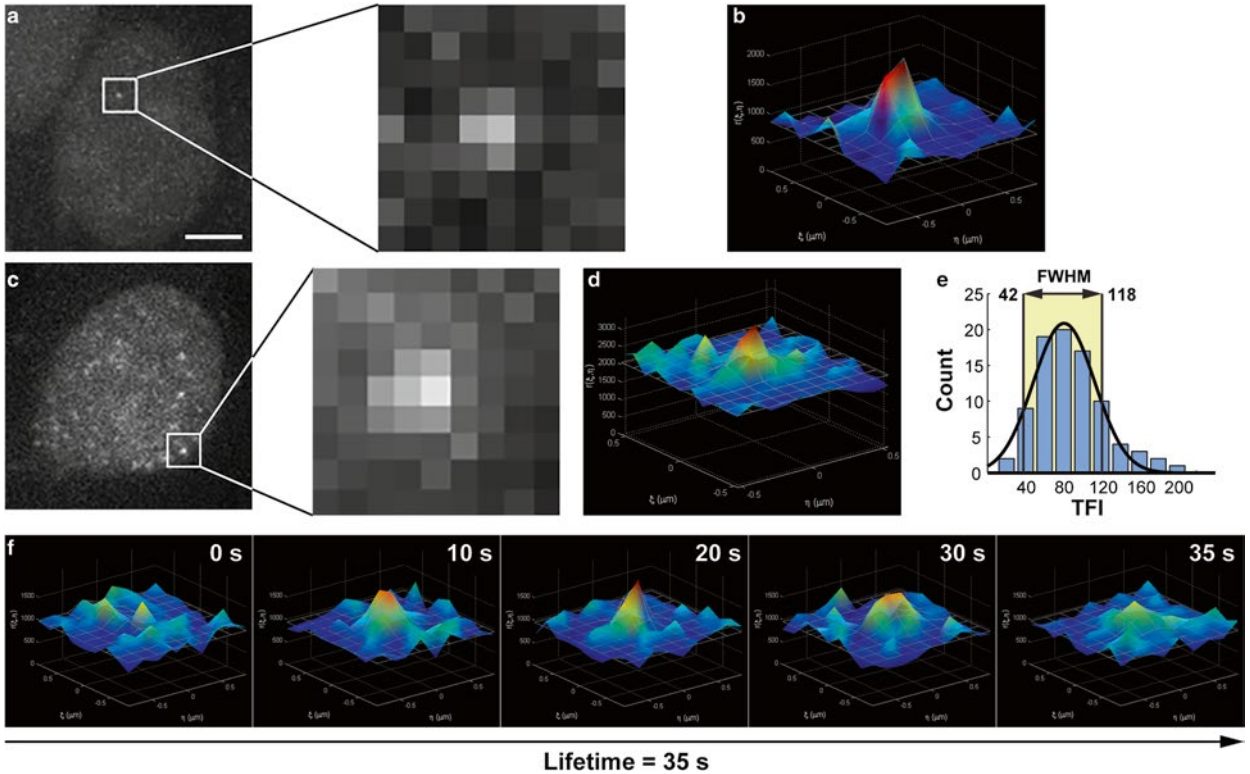

Fig. 3 Cell imaging and data analysis. (**a**) HEK 293 cells containing a single-β-globin reporter gene integrated in the recombination site of Flp-In T-Rex. The reporter gene is tagged with a cassette of 24 MS2 binding sites inserted in the second intron. Cells were imaged after transient transfection with a plasmid encoding MS2 coat protein fused in frame to GFP and a nuclear localization signal. The transcription site is detected as a fluorescent diffraction limited object. *Inset* shows higher magnification of the indicated region. (**b**) The total fluorescence intensity (TFI) at the transcription site is quantitated by Gaussian fitting. (**c**) HEK 293 cells containing a single β-globin reporter gene integrated in the recombination site of Flp-In T-Rex. The reporter gene is tagged with a cassette of 24 MS2-binding sites inserted in the terminal exon. Cells were imaged after transient transfection with a plasmid encoding MS2 coat protein fused in frame to GFP and a nuclear localization signal. Multiple fluorescent diffraction limited objects are detected diffusing in the nucleus. These objects correspond to single mRNAs in transit to the cytoplasm. *Inset* shows higher magnification of the indicated region. (**d**) The total fluorescence intensity (TFI) of single mRNA molecules is quantitated by Gaussian fitting. (**e**) Histogram depicting TFI values of single-mRNA molecules in the nucleus. The distribution was fitted to a Gaussian curve and the full width at half-maximum was calculated to estimate the range of TFI values corresponding to a single transcript. This range is then applied to time traces to select cycles of fluorescence gain and loss corresponding to transcription and splicing of a single transcript. (**f**) Intron lifetime is directly estimated as the time between fluorescence increase above background and its return to background levels

mRNA molecule that was released from the site of transcription and is in transit to the cytoplasm.

11. In case the fluorescent tag is intronic, select for a transcription dot with good signal to noise ratio and proceed to 4D imaging. Use the focus control to adjust the focal plane that corresponds to the brightest image of the dot. Define this plane as the central focal plane. Define a range of 3–4 planes above and below the central focal plane. Define also the distance in the z-axis between planes (*see* **Note 6** and Table 1). Acquire a *z*-stack of images every 5 s for a period of 5–10 min.
12. In case the fluorescent tag is inserted in the terminal exon, select a focal plane with multiple nuclear dots with good signal to noise ratio and acquire a single image.

3.4 Image Analysis

3.4.1 Image Format Conversion

1. From the Slidebook software, data are exported as 16 bit TIFF files. The conversion will save the exported data as a single file per acquired image or 4D series with the dimension order xyzt. This conversion is necessary to be able to use the data in image analysis software packages such as ImageJ (http://imagej.nih.gov/ij/) or MatLab (The MathWorks, Inc. https://www.mathworks.com/products/matlab/).

3.4.2 Quantitative Analysis of Intron Dynamics at the Site of Transcription

1. Intron dynamics at the site of transcription is analyzed from time lapse sequences composed of stacks of optical sections centered on the transcription site recorded at 5-s intervals.
2. The *xy* position of the transcription site in the first frame of the sequence is determined as a local intensity maximum, and a volume of interest is defined based on the acquired *z*-stack. The *z* plane corresponding to the highest fluorescence intensity value at the *xy* coordinates of the transcription site is recorded as its *z* position.
3. The *xyz* coordinates of the transcription site for subsequent time points are automatically determined by re-centering the volume of interest at the site for each time point. This is done by searching for a local intensity maximum, provided that the signal-to-noise ratio of fluorescence in the volume of interest is higher than a given threshold (typically 6).
4. The total fluorescence intensity (TFI) of the transcription site is calculated for each time point by performing a 2D Gaussian fit on the volume of interest at the *z* plane corresponding to the highest intensity value (Fig. 3b, d), with a modified implementation of the Gaussian fitting function developed by David Kolin (http://www.cellmigration-gateway.com/resource/imaging/icsmatlab/ICSTutorial.html) using the formula

$$I(x,y) = A + I_0\, e^{-\frac{\left(\frac{x-x_0}{a}\right)^2+\left(\frac{y-y_0}{b}\right)^2}{2}}$$

where A is the background nuclear fluorescence intensity, $I0$ is the peak intensity, $x0$ and $y0$ are the coordinates of the peak, a is the Gaussian width in x, and b is the Gaussian width in y. All these parameters are estimated by the fitting function. The total fluorescence intensity of the transcription site is calculated as the integral of the Gaussian curve: $TFI = 2\pi abI_0$ [21]. For transcription sites in which the difference between the maximum intensity value $I0$ and the background A is smaller than the standard deviation of the background σA estimated at the edges of the volume of interest (i.e., $|I0 - A| < \sigma A$), the TFI is set to zero.

5. Transcription sites for which the z plane of highest fluorescence intensity corresponded either to the first or last planes of the z-stack are discarded from the analysis.
6. The estimated TFI values for each transcription site are plotted over time. Time traces show fluctuations in fluorescence intensity, with periods of increment in fluorescence followed by periods of fluorescence loss. Increments in fluorescence intensity are due to de novo synthesis of MS2-binding sites immediately followed by binding of fluorescent MS2 fusion proteins, whereas fluorescence loss can result from either splicing followed by rapid degradation of the excised intron, or release of unspliced transcripts from the site of transcription. Determining whether splicing of the reporter gene occurs predominantly co- or posttranscriptionally is critical for the interpretation of fluorescence fluctuations (*see* **Note 7**).

3.4.3 Calibration for Single Transcripts

1. Multiple fluorescent transcripts can be simultaneous present at the transcription site. To count the number of RNA molecules imaged at the transcription site, the TFI of individual transcripts is estimated. This is achieved by measuring the TFI of mature mRNAs labeled in the terminal exon that are released from the transcription site and diffuse in the nucleus (Fig. 3c).
2. From cells that express the fluorescent tag inserted in the terminal exon, select a focal plane with multiple nuclear dots with good signal-to-noise ratio and acquire a single image. Acquire images from 20 to 30 different cells in at least three independent experiments. In each image, each individual spot is fitted by a 2D Gaussian function, the integral of which yields the Total Fluorescence Intensity (TFI) value (Fig. 3d).
3. Plot a histogram depicting TFI values of individual mRNA molecules diffusing in the nucleus. A single-peaked distribution of fluorescence intensity values indicates a homogeneous population of mRNA molecules (Fig. 3e).
4. Fit the distribution to a Gaussian curve. Calculate the full width at half-maximum (FWHM) to estimate the range of most frequent TFI values that correspond to a single transcript (Fig. 3e).

3.4.4 Direct Measurement of Intron Lifetime

In the case of co-transcriptional splicing, loss of fluorescence at the transcription site reflects intron excision. In this case, we define intron lifetime as the time from transcription of the MS2-binding sites to intron excision and degradation. We directly measure intron lifetime as the duration of cycles of fluorescence gain and fluorescence loss (i.e., starting and ending at fluorescence background levels) that involve synthesis of a single transcript. Cycles involving synthesis of a single transcript are defined as those that reach a maximum within the range of TFI values estimated in Calibration for Single Transcripts (Fig. 3f).

4 Notes

1. The sequence encompassing 24 copies of MS2-binding sites has 1093 bp, which corresponds to a linear length of approximately 372 nm. Because the sequence of each MS2-binding site forms a stem loop, the actual length of the fluorescently labeled sequence is roughly reduced by half. According to Abbe's formula and Rayleigh criterion ($d = 0.61 \times \lambda / \text{N.A.}$), at peak GFP emission wavelength of 509 nm and using a 1.4 N.A. objective, the lateral resolution limit is 222 nm. Thus, the size of the fluorescence source in each RNA molecule is below the resolution limit.
2. The MS2-GFP and MS2-mCherry fusion proteins contain a nuclear localization signal. The fusion proteins therefore appear confined to the nucleus.
3. Commonly used devices for live-cell imaging of gene expression are custom build wide-field fluorescence microscopes equipped with cooled CCD cameras. An alternative system that is widely used to analyze intracellular dynamics of single molecules with high spatial and temporal resolution is spinning disk confocal microscopy, which combines high sensitivity with high-speed optical sectioning and minimal photobleaching.
4. The intronic integration of the sequence encoding the MS2 stem loop array should be at least 100 bp away from the 5' splice site and branch point sequence in order to minimize the risk of disturbing the splicing reaction. In the case of exonic integration, it should be in the 3' UTR in order to avoid activation of NMD due to introduction of premature termination codons.
5. We recommend preparing dishes with different amounts of cells seeded. The optimal density of cells for transfection prior to an imaging experiment corresponds roughly to 60 % confluence.

6. The optimal distance between optical planes (z-step size) is estimated according to the Nyquist theorem, to minimize loss of information during optical *z*-sectioning.
7. To determine whether introns tagged with the MS2 system are spliced co-transcriptionally (i.e., before release of the mRNA from the transcription site), we use a variety of experimental approaches. First, RNA isolated from cells co-expressing the MS2-tagged reporter gene and fluorescent MS2 protein is analyzed by RT-PCR. The RNA is reversed transcribed using an oligonucleotide that is complementary to a sequence downstream of the transgene poly(A) site; the resulting cDNA is then PCR amplified using primers that specifically detect spliced and unspliced transcripts. This RT-PCR experiment detects transcripts that have not yet been cleaved at the poly(A) site and are therefore still associated with the gene template. We observed that uncleaved transcripts are predominantly spliced, arguing for co-transcriptional splicing [12]. Second, we reason that if mRNAs were released unspliced, spots should be detected emanating from the transcription site. Consistent with this view, live-cell imaging of cells treated with spliceostatin A, a potent splicing inhibitor, revealed a multitude of diffraction-limited objects diffusing in the nucleus [12]. We therefore conclude that if MS2-tagged introns are exclusively detected at the transcription site, it is most likely that they are co-transcriptionally spliced. Third, the dynamics of two introns in the same pre-mRNA was simultaneously visualized by double-labeling experiments. If transcripts were released unspliced, then fluorescence associated with the first intron should increase before fluorescence associated with the second intron and both fluorescent signals should decrease simultaneously. Rather, we observed that the first intron was excised while the second intron was still present in the nascent transcript, arguing against release of unspliced transcripts [12].

Acknowledgements

We gratefully acknowledge Tomas Kirchhausen and members of the Kirchhausen lab for advice and support during the development of this protocol. This work was supported by Fundação para a Ciência e Tecnologia, Portugal (PTDC/SAU-GMG/118180/2010; SFRH/BPD/66611/2009), and the Harvard Medical School-Portugal Program in Translational Research and Information.

References

1. Ellinger PHA (1929) Mikroskopische beobachtungen an lebenden organen mit demonstrationen (Intravitalmikroskopie). Arch Exp Pathol Pharmak 147
2. Heimstädt O (1911) Das fluoreszenzmikroskop. Z Wiss Mikrosk 28:330–337
3. Ploem JS (1967) The use of a vertical illuminator with interchangeable dichroic mirrors for fluorescence microscopy with incidental light. Zeitschrift fur wissenschaftliche Mikroskopie und mikroskopische Technik 68:129–142
4. Rino J, Braga J, Henriques R, Carmo-Fonseca M (2009) Frontiers in fluorescence microscopy. Int J Dev Biol 53:1569–1579
5. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. Science 263:802–805
6. Heim R, Prasher DC, Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci U S A 91:12501–12504
7. Shimomura O, Johnson FH, Saiga Y (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. J Cell Comp Physiol 59:223–239
8. Robinett CC, Straight A, Li G, Willhelm C, Sudlow G, Murray A, Belmont AS (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol 135:1685–1700
9. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2:437–445
10. Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, Prasanth KV, Ried T, Shav-Tal Y, Bertrand E, Singer RH, Spector DL (2004) From silencing to gene expression: real-time analysis in single cells. Cell 116:683–698
11. Shav-Tal Y, Darzacq X, Shenoy SM, Fusco D, Janicki SM, Spector DL, Singer RH (2004) Dynamics of single mRNPs in nuclei of living cells. Science 304:1797–1800
12. Martin RM, Rino J, Carvalho C, Kirchhausen T, Carmo-Fonseca M (2013) Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep 4:1144–1155
13. Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemin JP, Auboeuf D, Aitken S, Bertrand E (2011) Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. J Cell Biol 193:819–829
14. Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR (2014) Kinetic competition during the transcription cycle results in stochastic RNA processing. eLife 3
15. O'Gorman S, Fox DT, Wahl GM (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251:1351–1355
16. Sauer B (1994) Site-specific recombination: developments and applications. Curr Opin Biotechnol 5:521–527
17. Johansson HE, Dertinger D, LeCuyer KA, Behlen LS, Greef CH, Uhlenbeck OC (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. Proc Natl Acad Sci U S A 95:9244–9249
18. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH, Bertrand E (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. Curr Biol 13:161–167
19. Hocine S, Raymond P, Zenklusen D, Chao JA, Singer RH (2013) Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nat Methods 10:119–121
20. Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A 89:5547–5551
21. Rust MJ, Bates M, Zhuang X (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods 3:793–795

Index

Erik Dassi (ed.), *Post-Transcriptional Gene Regulation*, Methods in Molecular Biology, vol. 1358,
DOI 10.1007/978-1-4939-3067-8, © Springer Science+Business Media New York 2016

N

O

P

Q

R

S

T

U

Y