Advances in Experimental Medicine and Biology 907

Gene W. Yeo *Editor*

RNA Processing

Disease and Genome-wide Probing

Springer

Advances in Experimental Medicine and Biology

Volume 907

More information about this series at http://www.springer.com/series/5584

Gene W. Yeo
Editor

RNA Processing

Disease and Genome-wide Probing

Springer

Editor
Gene W. Yeo, Ph.D., M.B.A.
Department of Cellular and Molecular Medicine
Institute for Genomic Medicine
Stem Cell Program
University of California, San Diego
La Jolla, CA, USA

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-319-29071-3 ISBN 978-3-319-29073-7 (eBook)
DOI 10.1007/978-3-319-29073-7

Library of Congress Control Number: 2016936562

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

Ribonucleic acid (RNA) binding proteins currently number in the thousands, and defects in their function are at the heart of diseases such as cancer and neurodegeneration. RNA-binding proteins have become implicated in the intricate control of surprisingly diverse biological settings, such as circadian rhythm, stem cell self-renewal, oncogenesis, and germ cell development.

RNA-binding proteins bind to diverse RNA substrates to mediate aspects of the life cycle of RNA processing, including alternative splicing, stability, export, transport, and translation, as well as the assembly and disassembly of RNA granules. Understanding the roles of RNA-binding proteins requires multidisciplinary approaches: characterization of their targets, single-molecule approaches to unveiling their modes of interaction, and computational considerations of the tremendous amounts of data associated with analysis of their activities.

Given the success of the volume *Systems Biology of RNA Binding Proteins*, I have assembled this second volume with chapters focused on cutting-edge methods to study aspects of RNA-binding protein-RNA function that were not previously covered and introduce the most recent exciting biology related to RNA-binding proteins.

The content of this book surveys a range of genome-wide and systems approaches to studying RNA-binding proteins and the importance of RNA-binding proteins in development, cancer, and circadian rhythm. These chapters provide opportunities for open questions and new areas of inquiry into posttranscriptional processing.

With the combination of high-throughput short-read sequencing with biochemical methods such as RNA immunoprecipitation and RNA cross-linking and immunoprecipitation, transcriptome-wide maps of sites of interaction between RNA-binding proteins and their RNA targets have proved useful in elucidating novel molecular and cellular functions of RNA-binding proteins. Eric Van Nostrand and Stephanie Huelga in my laboratory discuss challenges associated with large-scale identification and analysis of RNA-binding protein-RNA interactions.

RNA does not exist as simple linear sequences in vivo. Gregory and colleagues present a review of genome-wide approaches to determining RNA structure in the

transcriptome. RNA structure discovery is a fundamentally important aspect of understanding the protein-RNA landscape.

With the increasing number of proteins thought to have RNA-binding capacity, Tomas Bos, Julia Nussbacher, and Stefan Aigner in my laboratory review how tethered function assays can be utilized to reveal novel molecular functions of candidate RNA-binding proteins.

To truly understand the stepwise interactions of proteins onto RNA substrates, Moore and Serebrov discuss single-molecule approaches that shed light on deep mechanistic insights into interrogating RNA-binding protein function.

The following chapters deal with exciting areas in biology that have become entwined with RNA-binding proteins.

Circadian control has recently become associated with posttranscriptional control of gene expression. Panda and colleagues recount circadian control of RNA by RNA-binding proteins.

Posttranscriptional and translation regulation is particularly important in germ cell biology. Licatalosi summarizes the role of RNA-binding proteins in male germ cell development.

Ito and colleagues present the importance of RNA-binding proteins in stem cell biology and oncogenesis. Self-renewal of stem cells is key to our understanding of development and cancer. How RNA-binding proteins control stem cell fate and oncogenesis is a topic of hot debate.

Hundley and Washburn feature the roles of RNA-binding proteins in RNA editing. Uncovering how enzymes affect RNA editing is critically important in our dissection of human diseases. It is still unknown how many and which RNA-binding proteins can regulate A-to-I editing in vivo.

Bejar clarifies the recent findings that splicing factors are commonly mutated in cancer. Recurrent somatic mutations of genes encoding core subunits of the spliceosome have been identified in several different cancer types and will be discussed in this review, avenues for novel cancer therapeutic strategies.

Calarco and colleagues discuss how tissue-specific alternative splicing is controlled and the utility of the model organism *C. elegans* in aiding in its elucidation.

Leung and Fan present exciting findings of the roles of RNA-binding proteins in RNA granules, particularly in neurodegenerative diseases such as ALS and FTLD.

Massirer and colleagues emphasize the importance of posttranslational modifications on RNA-binding proteins, an area poorly understood but with tremendous potential for understanding diseases such as neurodegeneration.

The contributors of this book are internationally recognized leaders in the arena of technology development, RNA processing, and biology relevant to RNA-binding proteins, and we envision that this book will serve as a valuable resource for both experts and nonexperts. Advanced undergraduate students and entering graduate students in biology, chemistry, molecular engineering, computer science, and bioinformatics, as well as medical students and postdoctoral fellows who are new to the arena of posttranscriptional gene regulation, should find this book accessible. We hope the chapters in this volume will stimulate interest and appreciation of the complexity and importance of posttranscriptional gene regulation to its readers and

even lead them to pose new solutions to the enormous challenges that lie ahead in comprehending how RNA-binding proteins affect gene regulation.

I sincerely express my greatest gratitude to the contributors to this book: Stefan Aigner, Rafael Bejar, Giorgia Benegiamo, Mario Bengtson, Nathan Berkowitz, Tomas Bos, Steven Brown, Kristina Buac, John Calarco, Alexander Fan, Sager Gosai, Xicotencatl Gracida, Brian Gregory, Ayuna Hattori, Stephanie Huelga, Heather Hundley, Takahiro Ito, Anthony Leung, Donny Licatalosi, Michael Lovci, Katlin Massirer, Melissa Moore, Adam Norris, Julia Nussbacher, Satchidananda Panda, Victor Serebrov, Ian Silverman, Eric Van Nostrand, and Michael Washburn.

La Jolla, CA, USA Gene W. Yeo

Contents

Contributors

Stefan Aigner Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Rafael Bejar, M.D., Ph.D. Division of Hematology and Oncology, UC San Diego Moores Cancer Center, La Jolla, CA, USA

Giorgia Benegiamo Institute of Pharmacology and Toxicology, University of Zürich, Zürich, Switzerland

Salk Institute for Biological Studies, La Jolla, CA, USA

Mario H. Bengtson Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas, UNICAMP, Campinas, Sao Paulo, Brazil

Nathan D. Berkowitz Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

Tomas J. Bos Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Steven A. Brown Institute of Pharmacology and Toxicology, University of Zürich, Zürich, Switzerland

Kristina Buac Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

John A. Calarco FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA

Alexander C. Fan Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Sager J. Gosai Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

Xicotencatl Gracida FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA

Brian D. Gregory Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

Ayuna Hattori Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

Stephanie C. Huelga Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Heather A. Hundley Medical Sciences Program, Indiana University, Bloomington, IN, USA

Takahiro Ito Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA

Anthony K.L. Leung Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Donny D. Licatalosi Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH, USA

Michael T. Lovci Center for Molecular Biology and Genetic Engineering, University of Campinas, CBMEG-UNICAMP, Campinas, Sao Paulo, Brazil

Katlin B. Massirer Center for Molecular Biology and Genetic Engineering, University of Campinas, CBMEG-UNICAM, Campinas, Sao Paulo, Brazil

Melissa J. Moore, Ph.D. Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA, USA

Adam D. Norris FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA

Eric L. Van Nostrand Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Julia K. Nussbacher Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Satchidananda Panda Salk Institute for Biological Studies, La Jolla, CA, USA

Victor Serebrov, Ph.D. Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA, USA

Ian M. Silverman Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, USA

Michael C. Washburn Department of Biology, Indiana University, Bloomington, IN, USA

Gene W. Yeo Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Molecular Engineering Laboratory, A*STAR, Singapore, Singapore

Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Chapter 1
Experimental and Computational Considerations in the Study of RNA-Binding Protein-RNA Interactions

Eric L. Van Nostrand, Stephanie C. Huelga, and Gene W. Yeo

Keywords CLIP-seq • eCLIP • HITS-CLIP • PAR-CLIP • iCLIP • Ribo-seq • BRIC_seq • RNA-seq • Microarray • RNA-binding protein • RNA processing

1 Background

The simplicity of the central dogma of molecular biology that information encoded by DNA is transmitted via RNA to proteins (the essential building blocks of cells) masks the complex regulatory networks involved at each step of this process. In eukaryotes, transcribed RNA molecules undergo a number of modifications that include pre-mRNA splicing, nucleotide editing and polyadenylation, all of which are tightly coupled to their sub-cellular localization, presentation and accessibility to ribosomal proteins for translation. At each of these steps the RNA may interact with a combination of one or more of the hundreds to thousands of RNA-binding proteins (RBPs) present in humans, providing a means of regulating the fate of these RNAs in a tissue-, temporal-, or condition-specific manner [1].

The importance of RNA-binding proteins in controlling RNA processing has been exemplified by studies of individual RNA substrates for decades. Recent advances in microarray and high-throughput sequencing technologies have made it possible to not only identify individual RNA-protein interactions, but to identify

E.L. Van Nostrand • S.C. Huelga
Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

G.W. Yeo (✉)
Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Molecular Engineering Laboratory, A*STAR, Singapore, Singapore

Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
e-mail: geneyeo@ucsd.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_1

RNA-binding protein targets genome-wide in a single experiment. RNA targets can be directly bound and also regulated by the RBPs, bound but unaffected by the association of the RBP, or not bound but indirectly affected by the RBP (Fig. 1.1). In this chapter, we will discuss recent efforts to both identify and predict *in vitro* and *in vivo* RNA-binding sites, and to integrate these direct targets with transcriptome profiling experiments in order to obtain a full picture of RNA regulatory networks. In particular, we will discuss the computational challenges implicit in expanding from analysis of targets of a single RBP to integration of target information for dozens or hundreds of the estimated thousand RBPs in the human genome.

2 What Is an RNA-Binding Protein?

Indeed, the question of how many RNA-binding proteins there are remains an open and challenging area of research. Most commonly studied RNA proteins interact with RNA through well-characterized RNA-binding domains, including zinc finger (ZNF), RNA recognition motif (RRM), and helicase domains that can interact with RNA in various forms (e.g. unstructured, single-stranded, double-stranded or a combination of both). Using such protein domains to search the human genome yields estimates of ~600 RNA-binding proteins [1]. However, various examples of RNA-binding proteins with non-canonical interactions have long suggested that other modes of interactions remain uncharacterized, suggesting that the number of true RNA-binding proteins may be larger than expected.

Recently, efforts to identify novel RNA-binding proteins using various approaches have further expanded the set of putative RNA regulators. In yeast, protein microarrays (in which a large number of proteins are spotted on a microarray and then probed with specific RNA transcripts of interest) identified a significant number of previously unannotated RNA-binding proteins, including surprising interactions between a number of enzymes and RNA [2, 3]. More recently, an RNA pull-down followed by mass spectrometry approach in human cells identified nearly 900 proteins interacting with RNA in human HeLa cells and 555 in mouse embryonic stem cells, including more than 300 and 250 respectively that were not previously annotated as RNA-binding proteins [4, 5].

In addition to raising interesting computational questions in the area of predicting which proteins will interact with RNA and with what mechanism, these results raise the question of what truly defines an RNA-binding protein. The single-stranded nature of RNA allows an RNA molecule to form a wide variety of secondary, tertiary, and even quaternary structures that enable a protein to not only interact with the ribonucleotide sequence, but also the nucleotide backbone in unique forms governed by the structure of the RNA. Indeed, RNA aptamers can be evolved *in vitro* that strongly interact with small molecules such as fluorophores [6] or specific proteins of interest [7, 8].

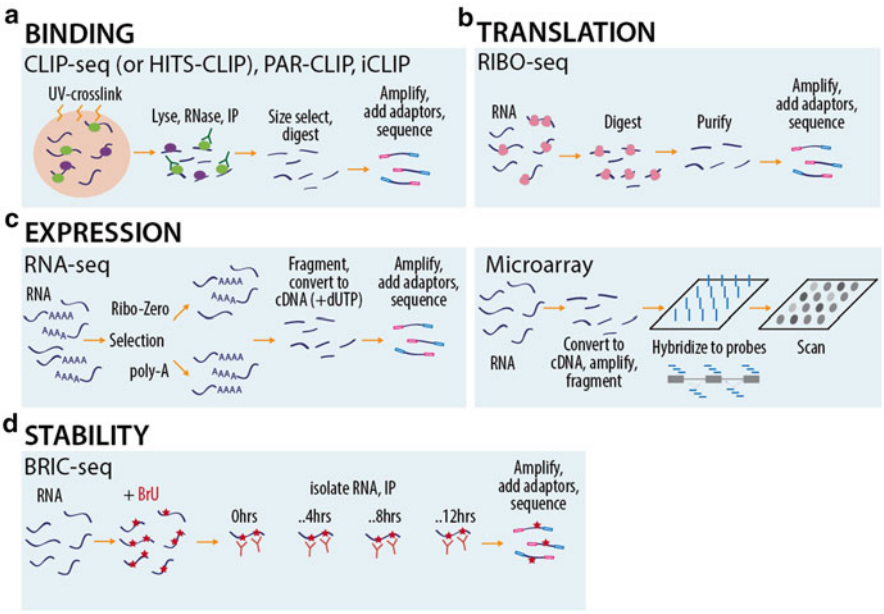


Fig. 1.1 Overview of methods to identify RNA-binding protein regulation of various RNA processing steps. (**a**) RNA molecules bound by RNA-binding proteins can be identified by a variety of related approaches generally termed CLIP-seq. Cell lysates are crosslinked (often with UV, to selectively cross-link protein-RNA interactions), and limited RNAse digestion is performed to generate small protein-bound RNA fragments. A RBP-specific antibody is then used to immunoprecipitate the protein of interest, along with associated RNA. After protein digestion, linkers are ligated to the 5′ and 3′ end of RNA fragments, which are then used to reverse transcribe and PCR amplify a DNA library for high throughput sequencing. (**b**) mRNAs being actively translated are identified using ribosome protection assays. Using the ART-seq kit (available from Epicentre), ribosome protected fragments are isolated by nuclease digestion followed by monosome purification. RNA is then purified, ligated to adapter sequences, and amplified for high-throughput sequencing. (**c**) Transcriptome-wide RNA expression and alternative splicing can be queried by RNA-seq (*left*) or microarray (*right*). For RNA-seq (*left*), ribosomal RNA is depleted using alternative methods, RNA is fragmented and reverse transcribed into cDNA, sequencing adapters are added, and then amplified for high-throughput sequencing. Stranded RNA-seq libraries can be generated by incorporating dUTP during second strand cDNA synthesis, followed by cleavage by UDG enzymes. For microarrays (*right*), RNA is converted to cDNA, amplified, fragmented, and fluorescently labeled. Labeled DNA is then hybridized to the microarray, and expression is read out by fluorescence intensity. (**d**) Other aspects of RNA processing can be studied using specialized protocols, such as BRIC-seq-based quantitation of RNA half-lives. After growth in media containing 5′-bromo-uridine (BrU), fresh media is added to terminate labeling of newly synthesized RNAs with BrU. Pull-down of remaining RNA with an anti-BrU antibody at various timepoints, followed by library preparation using standard RNA-seq protocols, can then yield quantitative estimates of RNA stability

If an RNA can be engineered to specifically bind to GFP with high affinity, does this mean that GFP is an RNA-binding protein? Conversely, if an RNA- or protein-pulldown indicates an *in vivo* RNA-protein interaction, that would appear to be convincing evidence that the protein is an "RNA-binding" protein. However, the term "RNA-binding protein" has traditionally been reserved for describing proteins that functionally interact with RNA—i.e., that the protein's interaction with RNA molecules causes some differential regulation of RNA processing in a regulated manner. Thus, as our abilities to detect *in vivo* protein-RNA interactions continue to improve, consideration will have to be given as to whether identification of RNA-protein interactions is sufficient, or whether it is more relevant to combine these results with analyses that also address whether this interaction with RNA drives the fate of the RNA.

3 Identification of RNA-Binding Protein Binding Sites *In Vivo*

A genome-wide view of RNA-binding protein interactions is essential to understand how RNA-binding proteins recognize RNA and drive differential RNA processing. These methods are generally referred to as "cross-linking followed by immunoprecipitation and high-throughput sequencing" or CLIP-seq, as a parallel to the ChIP-seq approaches used to purify DNA binding proteins and the associated bound chromatin (Fig. 1.2a). Many variations (HITS-CLIP, PAR-CLIP, iCLIP) have been described [9–12] (Fig. 1.2b), all of which follow the same basic strategy: crosslinking of protein-RNA interactions, digestion or shearing of the RNA to smaller fragments, immunoprecipitation of the targeted protein (and associated RNA) using a target-specific antibody, isolation of RNA, and ligation of linkers and RT-PCR amplification to generate libraries compatible with either microarrays or, more recently, high-throughput sequencing.

Various algorithms have been developed to identify regions of significant association from CLIP-seq experiments. Fundamentally, these algorithms attempt to solve the question of whether the read density within a given region is characteristic of a binding site or characteristic of random association. Early methods simply counted the number of reads within a defined window size (e.g., 100 nucleotides) and asked whether the number of reads was significantly enriched above expected, given the number of total reads throughout that transcript. More recent methods have incorporated additional features, such as improved statistical modeling of read distributions, information about repetitive elements and read mappability, and the canonical 'shape' of a binding site (the local distribution of reads around a binding site) to improve identification of true binding sites [9, 13].

Depending on the experimental methodology used, other information can also be incorporated into binding site identification. Using general CLIP-seq techniques, typical cluster sizes are ~50–150 wide, depending on the size selection used during library preparation. Recognizing that the reverse transcription step

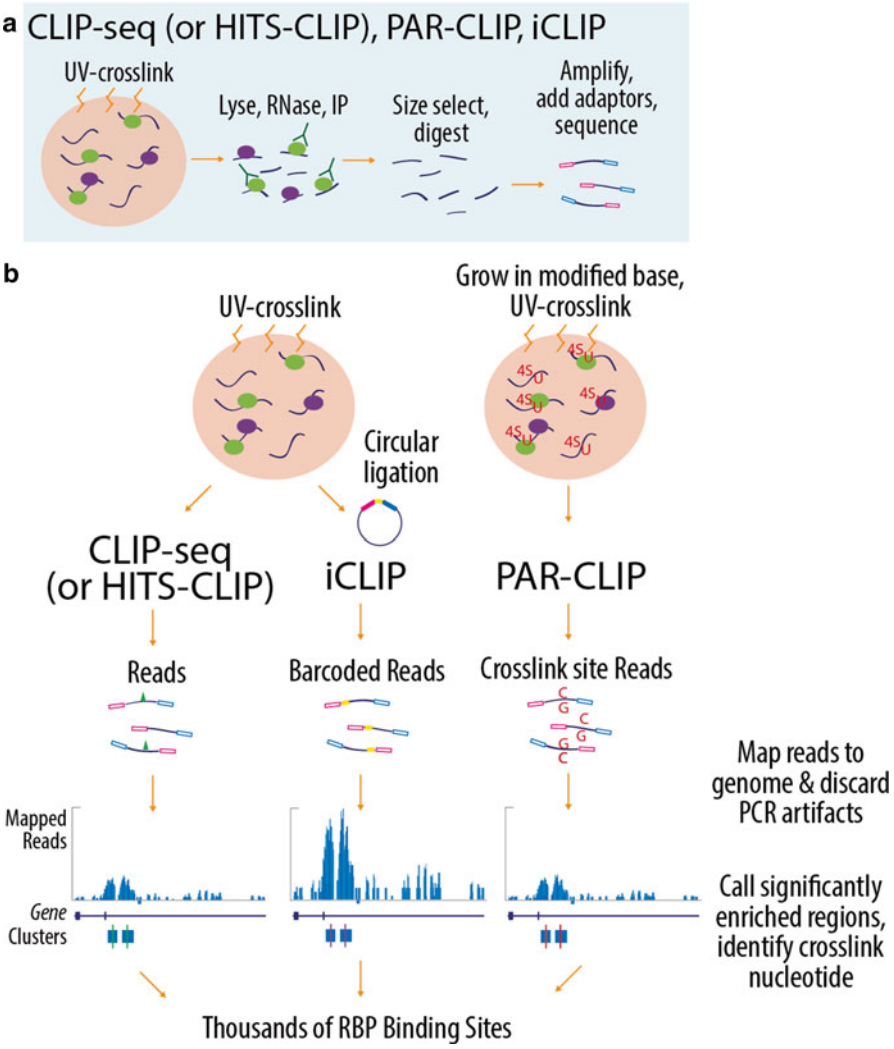


Fig. 1.2 Identification of direct RNA-binding protein targets by CLIP-seq. (**a**) Overview of CLIP-seq procedure (repeated from Fig. 1.1). (**b**) Several variants of the general CLIP-seq method (*left*) have been developed to optimize various steps. iCLIP (*center*) uses a single linker oligonucleotide coupled with a circular ligase enzyme to improve linker ligation efficiency [9]. This linker also adds a barcode as well as five random nucleotides to each sequencing read, allowing one to disambiguate PCR amplification artifacts from truly independent RNA fragments, increasing the dynamic range of sequencing read depth. In PAR-CLIP (*right*), photo-reactive ribonucleotide analogs are incorporated into transcribed RNA in order improve crosslinking efficiency [10]. In addition, the cross-linked uridine will be complemented by a guanine base in the reverse transcribed cDNA during library preparation step. This characteristic U to C mutation in the center of a binding site can be used to identify binding at single-nucleotide resolution. Single-nucleotide resolution binding sites can also be computationally identified from normal CLIP-seq (where reverse transcription often skips the cross-linked nucleotide, leading to single-base deletions (*green triangle*) in the sequencing reads), or iCLIP (where reverse transcription often terminates at the crosslinked base pair, leading to read pileups at the crosslink site)

was often inefficient and terminated at the nucleotide cross-linked to the RNA-binding protein, Konig *et al.* [9] exploited this truncation to anneal a linker oligonucleotide at this terminal end. Termed 'iCLIP', this approach led to a pile-up of reads at the cross-link site, enabling single-nucleotide resolution identification of binding sites [9]. Using more traditional CLIP datasets, Zhang & Darnell recognized that single-nucleotide resolution could also be achieved computationally by searching sequencing reads for single nucleotide insertions and deletions characteristic of reverse transcriptase skipping at the cross-link site [14]. Although less than 20 % of reads showed such mutations, this density was sufficient both to identify the bound residues at most binding sites, as well as to generate insights into binding motifs and local sequence structure at binding sites genome-wide [14].

Before cluster identification, an important step is quality control to determine whether the CLIP-seq experiment was successful and generated reliable data. As immunoprecipitation of RNA-binding proteins (and subsequent adapter ligation onto RNA) is often inefficient, CLIP-seq libraries are often amplified from small amounts of RNA input, leading to significant concerns about PCR amplification biases when a high number of PCR cycles are needed to amplify enough DNA for sequencing. Indeed, a high degree of read redundancy (multiple reads of identical sequence) is typically observed in CLIP-seq experiments, up to 90 % in many published datasets [15]. However, as the identified clusters are typically short (<200 nt) and (as described above) often start at the cross-linked nucleotide, it is impossible to distinguish PCR amplification from multiple unique fragments using the standard CLIP methodology, and most approaches will simply compress these multiple reads into one for downstream analysis. Paired-end sequencing can address some of these concerns (as the odds are lower that two unique fragments would have the same start and end position). iCLIP provides further ability to distinguish these possibilities by incorporating a random five base sequence (along with a multiplexing barcode) into the ligated RNA adapter that will ultimately be sequenced in each read [9]. After sequencing, computational processing can be performed to identify the random adapter sequence, barcode, and RNA fragment sequence. Using this approach 1024 (4^5) reads can map to the same location instead of only one, significantly improving signal to noise at true binding sites.

This high inefficiency in library generation led us to develop an enhanced CLIP (eCLIP) methodology for large-scale CLIP experiments performed as part of the ENCODE consortium efforts [15]. By modifying the adapter ligation steps to 70–90 % efficiency, the rate of PCR duplication was decreased by ~60 % with a concomitant decrease in experimental failure rate. These improvements enabled the successful performance of 102 eCLIP experiments profiling 73 RNA-binding proteins in HepG2 and K562 cell lines in biological duplicate [15]. Additionally, these improvements enabled generation of a paired size-matched input for normalization, which substantially improves signal-to-noise in identifying true peaks by enabling removal of common false positive signals [15]. This and other advances in CLIP methodologies should rapidly advance our ability to identify RNA-binding protein binding sites in vivo at large scale.

4 Challenges of Peak Finding for RNA-Binding Proteins Compared to DNA Binding Proteins

It is appealing to turn to the field of DNA binding proteins, as the identification of *in vivo* binding sites for DNA-binding transcription factors requires solving a number of similar computational challenges. Akin to identifying RNA-binding protein binding sites, the experimental association of transcription factors to their DNA binding sites involves cross-linking the factor to its genomic DNA (typically using formaldehyde), fragmenting the DNA, immunoprecipitating the factor and isolating the associated 100–150 bp DNA fragments, and then quantifying the associated DNA using high-throughput sequencing (chromatin immunoprecipitation followed by sequencing, or ChIP-seq) [16].

Although the fact that it is RNA instead of DNA being queried seems like a trivial difference, in reality there are substantial challenges distinct from identifying DNA binding sites. In a ChIP-seq experiment in normal tissues, the expected read density across the genome is roughly equal (as all DNA regions are expected to be present on each of the two chromatids in each cell); although this genome-wide equivalence may not hold for all cell lines (which may have acquired chromosomal aberrations), the distribution in a reasonably large local region around the binding site should show this equivalency. In contrast, as each RNA transcript is expressed at a different level, all calculations to identify RNA-binding sites are limited to within-transcript read information. RNA splicing also represents a complication, as introns (which are among the most common binding sites for many RNA-binding proteins, and represent the functional binding sites for many splicing regulators) are rapidly spliced from most transcripts, meaning that only a small fraction of messages for a gene of interest will still contain a potential RNA-binding protein—intronic sequence interaction that can be quantified. Additionally, the presence of RNA requires additional, more complicated ligation steps to generate tagged RNA molecules that can be reverse transcribed into cDNA. Finally, because a transcription factor associates with double-stranded DNA, DNA binding sites obtained by ChIP-seq will be identified by sequencing reads that map to both the Watson and Crick strands. However, the distributions of these reads on the two strands will be separated by a width equal to the size to which the DNA is sheared, enabling this characteristic shift to be used in downstream algorithms as both a validation of true binding sites as well as a validation of proper library construction [17].

However, there are also more subtle differences. Although ChIP-seq experiments often can use a non-immunoprecipitated ‘input’ sample or an ‘IgG-only’ secondary antibody pull-down with no primary antibody as a negative control, those negative control samples have proven more difficult to obtain for CLIP-seq experiments. Due to the methodology of CLIP-seq, attempts to leave out the primary antibody have typically yielded too little RNA to yield useful high complexity libraries for sequencing. Because of this, most CLIP-seq analysis is still done using custom-written pipelines by individual labs, and has not yet coalesced into one generally

used software package across the field. However, recent improvements to CLIP methodologies have enabled the proper generation of paired size-matched input samples, which enable peak calling to be performed and validated at the level of enrichment over input background [15]. Future work to integrate standard ChIP-seq methodologies for both peak calling and validation (including tools to assay replication across biological replicate samples) should bring CLIP-seq analysis into a standardized framework for non-expert users.

5 RBP-Responsive RNA Targets

Identification of the *in vivo* binding sites for an RNA-binding protein can give insight into the molecular mechanism of binding for that factor. However, to understand the molecular basis for the phenotypes observed upon knockdown or overexpression it is necessary to identify not only those targets directly bound by the RBP, but also those that show RBP-dependent altered RNA processing. Individual target and deep mutational analysis of minigene-based assays have provided significant insights in the study of individual RNA-binding protein responsive targets over the past decades [18]. More recently, it has become possible to modulate an RNA-binding protein and query the effect on various outputs (RNA expression, alternative isoform usage, and ribosomal occupancy) in a genome-wide manner. However, inherent in each of these assays are various experimental decisions and computational challenges that should be considered during the initial planning stages.

6 Choosing Between Depletion Versus Over-Expression Experiments

The first decision when designing the experiment is whether to profile loss-of-function or gain-of-function of the protein of interest. It is ideal to mutate both the maternal and paternal alleles of the protein to generate a complete loss of function, as this ensures that the function of the protein is completely lost. However, creating such a null mutation can often be difficult and require significant investment of resources, particularly in whole organisms. More common is to knock down the protein, which can be performed using a variety of techniques that depend on the cell-type and duration of knockdown desired [19]. The commercial availability of knockdown reagents such as siRNA or shRNAs for nearly all protein-coding human and mouse genes, as well as the ease of generating and using these reagents, makes knockdown both the simplest as well as the most scalable experiment. However, there exist a number of caveats with knockdown experiments that are important to consider in downstream analyses. First, most knockdown reagents (including shRNA and siRNA molecules) use an antisense oligonucleotide which leads to RNA degradation through the RNA interference pathway, but which can have

uncharacterized off-target effects if the oligonucleotide binds to other expressed genes. As such, it is important to perform multiple independent knockdowns with antisense oligonucleotides targeting multiple regions of the gene, in order to separate true RBP targets from off-target effects [20]. Secondly, many RNA-binding proteins are found in families with multiple highly similar paralogs that may have functional redundancy. As such, knockdown of only one family member may yield few if any altered targets, which will only be revealed after simultaneous knockdown of multiple family members. Various degrees of redundancy have been observed for a variety of RNA-binding proteins, including the Muscleblind family [21] and RBFOX family [22, 23], indicating that this may represent a concern for analysis of many RBPs.

Ectopic expression of an RNA-binding protein can avoid this issue of functional redundancy, as well as avoiding concerns for off-target effects. Additionally, it is particularly powerful when the RBP of interest is introduced into a cell line in which it is not normally expressed, as this may help to isolate the signal of true targets of the RBP in isolation of other RBPs that it often interacts with. However, the expression of a transgene brings with it its own set of concerns. Significant over-expression of DNA- and RNA-binding proteins can lead to binding to (and subsequent regulation of) lower-affinity binding sites that would not be physiologically relevant *in vivo* targets [24]. Thus, it becomes important to design the promoter, 3′ UTR, and other aspects of the transgene in such a way to express the transgene at roughly equivalent levels to its expression in other tissues. This can also be addressed computationally, by comparing RBP-responsive targets against other information to verify whether the targets observed are likely to be relevant under normal conditions. Depending on the experimental method, over-expression may also require selection of an individual isoform of the RNA-binding protein if the entire gene region is too large to reliably transmit into the cell line of interest. As multiple isoforms of an RBP can show distinct regulatory activity and even differential subcellular localization [25], it becomes critical to consider whether the activity of the gene or only of that specific isoform is being read out.

7 Quantitation of RNA Isoform Abundance

The second major decision in identifying RNA-binding protein targets is what aspect of RNA processing will be queried, and how altered targets will be identified. Currently, much effort has been devoted to identification of alternative splicing events, which can be quantified using both high-throughput sequencing (RNA-seq) and microarray approaches. In both cases, the goal of these analyses is to quantify the percent of isoforms that contain an alternative exon (quantitated as the percent spliced in, or Psi (Ψ) value). However, the methodological and technical differences between these two technologies mean that the computational challenges for analysis, and the algorithms and approaches used to solve these challenges, are somewhat unique for each platform. Microarrays are best suited for quantitation of previously

identified alternative splicing events, whereas RNA-seq is able to identify alternative isoforms *de novo*. For both, there are two basic approaches to determine psi values: a direct approach, in which one typically looks for signal from the specific exon-exon junctions that correspond to various isoforms, or an indirect approach, where the level of each isoform is computationally inferred based on not only this direct information but also by modeling to predict inclusion in situations that lack direct evidence. Given infinite information, direct approaches are generally preferred; however, there are many cases where indirect approaches can either help gain insights with lower sequencing depth or can be used in cases where direct information is not possible.

8 Identification of Altered RNA Splicing Events by Microarray

Microarrays contain thousands of short oligonucleotides that are each complementary to a specific target sequence, ranging from 25-mers on Affymetrix platforms to 60-mers using Agilent arrays. These oligonucleotides are then spotted to a surface with a known pattern, and a fluorescently labeled cDNA library is flowed over the array surface. DNA fragments complementary to the spotted oligonucleotides will anneal at that position, and the fluorescence of each 'spot' can be used as a read-out for expression level of the sequence complementary to the oligonucleotide probe spotted at that position [26].

With development of high density microarrays, it is now common for the microarray to contain multiple probes that tile across a region of interest, providing multiple independent readouts of expression of an individual gene, exon, or splice junction (Fig. 1.3a–c). For quantitation of RNA processing, splice junction probes are particularly helpful as they can be used to directly measure expression of individual alternative isoforms. Thus, a cassette exon being queried will typically have six 'probe sets' of interest (each of which contain multiple individual probes): one probe set tiling the exon (represented as ASexon in Fig. 1.3b) as well as one tiling each of the upstream and downstream exons (UPexon and DNexon), and three splice junction probe sets tiling each of the three splice junctions (UPjunction, DNjunction, and ASjunction) (Fig. 1.3b). Quantitation of exon inclusion can then be performed by comparing the signal at probes supporting exon inclusion (probe sets ASexon, UPjunction, and DNjunction) against those supporting exon exclusion (probeset ASjunction) [27]. Quantification of other types of alternative splicing simply requires different probe positioning to measure the novel junctions generated by those events. Figure 1.3c shows representative data for SLK exon 13 in experiments profiling differentiation of induced pluripotent stem cells to motor neurons. Notably, while upstream and downstream constitutive exons showed stable expression between iPSC and differentiated neuron samples, significantly decreased signal is observed at all three probesets representing exon inclusion. Concordantly, signal was increased at the ASjunction probe that directly measures exon exclusion.

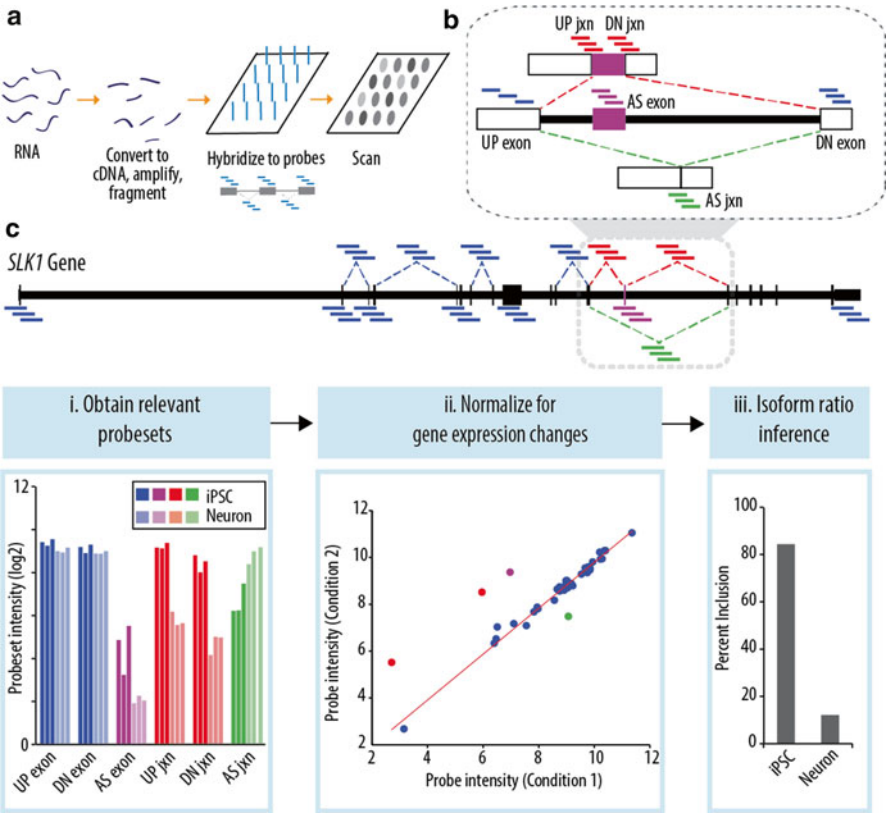


Fig. 1.3 Analysis of alternative RNA processing by microarray. (**a**) Overview of RNA expression quantification by microarray (repeated from Fig. 1.1). (**b**) Quantitation of alternative splicing events by microarray requires probes designed to target specific regions associated with the event. For a cassette exon, traditional exon microarrays will have probes assaying the desired alternatively spliced (AS) exon (*purple*) as well as flanking (UP and DN) constitutive exons (*blue*). More recent microarrays have incorporated additional probes assaying splice junctions (UP jxn and DN jxn) that quantify exon inclusion (*red*) or the alternatively spliced (AS) exclusion junction (*green*). (**c**) Splicing-sensitive microarrays identify alternative splicing of SLK exon 13 during neuronal differentiation. (*Left*) Although flanking exons show relatively similar signal intensity between control and knockdown samples, the three probe sets querying exon inclusion show ~8-fold higher signal in iPS cells as compared to differentiated neurons, whereas the probeset measuring the junction created by exon exclusion increases ~4-fold. (F.J.M. & G.Y. *unpublished data). (Center)* As a first step to identify significantly altered splicing events, change in probe intensity for splicing-sensitive probes are normalized for gene expression changes by regression of all other constitutive exon probes covering the SLK gene. (*Right*) Various algorithms have been described to convert the characteristic opposing changes in UPjxn/DNjxn and ASjxn probe intensity into an estimate of PSI value. These PSI values can then be validated by RT-PCR

Analysis of these microarrays can be challenging, with multiple methods described that incorporate these multiple probe sets into one value that characterizes differential processing between samples [27, 28]. In the simplest approach, intensity at probesets targeting neighboring constitutive exons can used to normalize for sample or gene expression differences, and performing a *t*-test between the three knockdown and three control replicate samples can reveal significantly altered individual exons [29]. The addition of probesets covering the splice junctions makes this calculation more robust by providing additional independent measurements that should all show the same fold-change between conditions [27, 28, 30, 31]. This has led to the development of probabilistic models that can infer the underlying expression of individual isoforms from probe intensity measurements (reviewed in *Chen* [32]).

9 Quantifying Alternative Splicing by High-Throughput Sequencing

Identification of alternative splicing by high-throughput sequencing (RNA-seq) is performed by starting with RNA from a sample of interest. As total RNA samples contain mostly ribosomal RNA, a purification to deplete rRNA is typically performed by either selection with poly-T oligonucleotides to isolate only polyadenylated RNA (enriching for mRNA), or by selective depletion of ribosomal RNA using bead-coupled antisense oligonucleotides (using kits such as the Ribo-Zero kit available from Illumina). RNA is then fragmented, reverse transcribed into cDNA using random hexamer primers, and the second cDNA strand is synthesized. Various methods can be used to preserve strand information; one commonly used technique involves replacing dTTP with dUTP in the second strand synthesis step, and then performing PCR with a polymerase enzyme that is blocked at dUTP nucleotides [33]. Oligonucleotides containing adapters for high-throughput sequencing are then ligated, and the sample is PCR amplified and size-selected for sequencing. Sequencing a sample using one lane of Illumina HiSeq 2500 or 4000 machines (the current methodology with greatest sequencing depth) yields over 200 million reads, each of which can be 30–250 bases and in either single-end or paired-end format (depending on user selection of sequencing kits). Reads are then mapped to the genome (typically using one of several publicly available programs [34, 35]), and post-processed along with a genome annotation to derive position- as well as transcript-level read density profiles [36].

Figure 1.4 shows the type of data utilized for two types of alternative isoforms: cassette exons (Fig. 1.4a), and alternative polyadenylation sites (Fig. 1.4b). For cassette exons, reads that map to the junction between the upstream and downstream exon provide strong, direct evidence of exon exclusion. In contrast, reads that map to the exon itself or that map to exon-exon junctions with the upstream or downstream exon provide evidence of exon inclusion. The ratio of inclusion reads (normalized for the increased mappable sequence length) relative to exclusion reads will

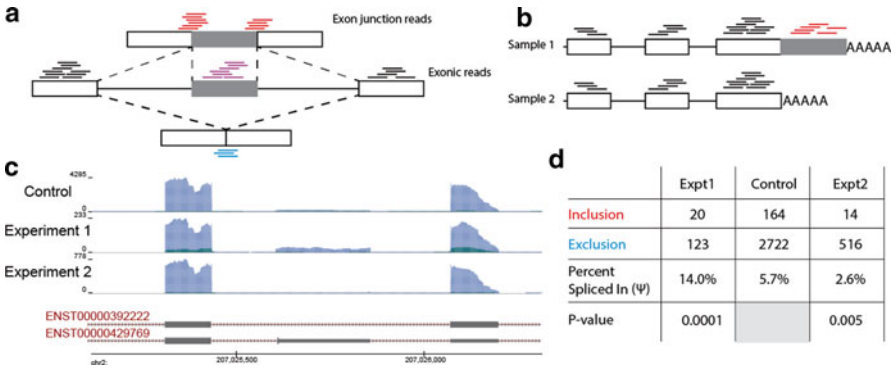


Fig. 1.4 Identification of altered RNA processing by RNA-seq. (**a**) Schematic representation of sequencing reads enabling quantitation and identification of an alternative cassette exon. Reads mapping to the *upstream* and *downstream* flanking exons (*black*) can be used to estimate the overall level of transcript abundance but are not informative for splicing of the internal exon. Reads mapping to the alternative exon (*purple*) or to the *upstream* or *downstream* splice junctions (*red*) provide evidence for exon inclusion, whereas reads mapping to the junction between the flanking exons (*blue*) provide evidence for exon exclusion. (**b**) Similar schematic for alternative polyadenylation site usage. Identification of alternative polyadenylation sites is more complex, as there are reads that uniquely map to the longer isoform (*red*), but all other reads are expected to map to both isoforms (*blue*). (**c**–**d**) Read density profile for exon 4 of EEF1B2 in three RNA-seq experiments, with height of the blue histogram indicating the number of reads that map with the 5′ end at the given base. The percent spliced in (Ψ) value is then defined as the number of reads supporting inclusion divided by the number supporting either inclusion or exclusion. Significance relative to a control sample can then be calculated by Fisher's Exact test (or the equivalent Chi-square test) on the number of inclusion- and exclusion-supporting reads

then provide a ratio of exon inclusion in a sample of interest (Fig. 1.4c). A simple hypergeometric or Pearson's Chi-square test between exclusion and inclusion reads can then be used to determine whether an event shows altered splicing across two samples (Fig. 1.4d) [37]. The simplicity of this approach, combined with the fact that all reads used in the analysis provide direct evidence either for an inclusion or exclusion isoform, make it ideal for experiments with extremely high sequencing depth (or for highly expressed transcripts). However, as exclusion can only be identified by reads that span the specific exclusion exon-exon junction, short read length will often lead to few reads that actually span this junction, limiting detection.

This direct quantitation of Psi values does not use any information about reads that map elsewhere in the transcript. However, expression of flanking exons can provide useful information—for example, if the upstream and downstream exons each show dramatically higher read densities than a central queried exon, this would suggest that the exon is likely excluded in some percent of transcripts. Although this drop in read density is only probabilistic evidence of an excluded exon and not direct observation of the excluded isoform, inclusion of such information provides more accurate estimates of exon inclusion rates [38]. Because isoform quantitation using this approach is now no longer deterministic (based on read counting) but

now requires estimation of the most likely Psi value that would lead to the observed read densities, algorithms that incorporate such information estimate these values using various machine learning and probabilistic approaches. MISO, one of the more popularly used software packages for this analysis, treats this estimation as a Bayesian inference and uses the observed read densities across the queried exons and exon-exon junctions as a 'posterior' that can be used to infer the 'prior', or true (but unobserved) Psi value [38]. The estimated Psi values, along with confidence intervals, can then be used to identify significantly altered alternative splicing between two tissues or conditions. Further development of other methods to robustly estimate Psi values should yield further improvements in quantitation of splicing changes by RNA-seq.

Because isoform quantitation by RNA-seq depends on sequencing reads that specifically map to the queried exon (and flanking exons and splice junctions), the total number of reads obtained from high-throughput sequencing represents a major consideration for analysis success. Detection of an alternative event thus depends on two factors in addition to the splicing event itself: overall sequencing depth, and expression of the entire gene. Splicing events in highly expressed genes will be easy to detect even with low-coverage experiments (for example, alternative intron retention in the 3′ UTR of NAP1L1 can be easily detected in less than ten million reads, achievable in a single run on an Illumina MiSeq machine). In contrast, alternative events in lowly expressed genes may be difficult to detect using more complex algorithms and the hundreds of millions of reads obtained in more deep sequencing on an Illumina HiSeq machine. Thus, analysis of RNA processing by RNA-seq will always involve a trade-off between sequencing depth (and thus, cost) and the ability to detect events of genes with lower expression (which may include biologically relevant events in DNA- or RNA-binding proteins that are not always highly abundant even when functional).

10 Identification of Novel Alternative Splicing Events by RNA-Seq and Microarray

Figures 1.3 and 1.4 describe how the expression of a previously identified alternatively spliced exon can be quantified. However, it is also possible to adapt these techniques for *de novo* identification of alternative isoforms, finding previously unseen isoforms that may be unique to specific cell-types or conditions. As described above, sequencing reads that perfectly map across a splice junction can be used to verify a novel alternative event; thus, identification of novel alternative events requires using sequencing reads that do not map to the annotated genome or transcriptome to infer the splicing event that must have occurred to generate the observed sequence. Although this is easiest with paired end reads, longer read length in high-throughput sequencing has improved the ability to detect novel splicing events using single-end reads as well.

Because of this additional step, algorithms that not only align reads but also perform *de novo* transcript assembly and isoform identification have typically been slower and required heavier computational resources than simple mapping approaches. The TopHat and Cufflinks suite of programs have been widely used to identify and quantify novel splicing events, but can require significant resources for large-scale datasets [36]. To handle the scale of RNA-seq data generated by the ENCODE consortium, the alignment program STAR was developed [34]. In STAR, the largest 5′ fragment of the read that maps to the genome is identified first. Then, the mapping is repeated for any remaining unaligned sequence, identifying split reads that identify splice junctions or other fusion products. Further processing can then identify the exons (whether annotated or novel) that these split reads are derived from. Although STAR requires considerable computational resources (~40gb of RAM for mapping to the human genome, although this can be relaxed to ~16gb at a trade-off to mapping speed), it enables rapid identification of novel alternative splicing events from large compendia of RNA-seq datasets [34]. Other algorithms use similar approaches to identify spliced reads, with various tradeoffs for speed and sensitivity.

Identification of putative alternative exons from microarrays is more challenging, but can be done in different ways depending on the microarray strategy employed. As described above, direct identification of an alternative isoform by microarray requires probes designed to span the exon-exon junction created by that alternative event. Thus, direct identification of novel alternative events requires tiling all exon-exon junctions (for example, probes that span the junction between exons 1 and 2, 1 and 3, 1 and 4, 1 and 5, etc.) in order to identify signals from previously unannotated splice events. Profiling of 52 tissues using such an approach identified thousands of novel alternative splicing events with a high accuracy rate [26]; however, such custom array designs are not commonly used due to cost, complexity of analysis, and the difficulty in balancing exhaustive coverage to identify novel events with wasted coverage of non-observed junctions. Indirect inference of novel splicing events can also be performed using more common exon-level microarrays, including the Affymetrix Human Exon arrays. As was the case with RNA-seq data, indirect identification of novel exon exclusion events can be inferred by comparing relative expression of an exon, as well as its upstream and downstream flanking exons, across multiple samples. However, this approach has lower detection ability and is limited to events that are assayed on the array, typically limiting analysis to simple exon skipping events.

Thus, RNA-seq and microarray based approaches each have independent advantages and disadvantages to identification and analysis of alternative RNA processing events. Microarrays can provide advantages to quantification of a known set of events, as all events can be easily quantified genome-wide in a single experiment. In contrast, the level of expression of each gene represents a significant hurdle in quantifying RNA processing by RNA-seq, as ever-increasing amounts of sequencing are required to detect alternative isoforms in lowly expressed genes. Recent methods of generating targeted RNA-seq libraries to enrich for specific desired sequences can alleviate this concern somewhat, but add significant cost to the

experiment. In contrast, RNA-seq is uniquely well-suited to identify novel alternative events, as it simply requires the development and implementation of computational tools to identify these events from a standard RNA-seq library. In contrast, identification of novel events by microarray requires either significant alteration to the design of the microarray itself, or specialized analysis tools that can detect exon skipping but are limited in their ability to detect more complex alternative splicing events. Thus, the methodology being chosen will depend heavily on the experimental design—RNA-seq is ideal when profiling a novel tissue or cell-type, or manipulation of an RNA processing factor that may have unknown roles, in order to identify novel and complex alternative events that may have been previously missed. In contrast, there remain some advantages to using microarray-based approaches for large-scale profiling of known sets of splicing events. However, continued improvements in high-throughput sequencing technologies (in read length, read number, and cost per read) make sequencing an ever-more appealing option not just for novel event discovery but also for robust quantitation of alternative splicing and RNA expression.

11 Alternative Polyadenylation Sites

Recent publications have identified alternative polyadenylation as a significant factor in functional isoform diversity, as the generation of isoforms that contain or lack an extended 3′ untranslated region can lead to alternative inclusion of microRNA target sites or other regulatory sequences that drive differential RNA regulation. Multiple groups have developed techniques to identify polyadenylation sites by RNA-seq, by identifying sequencing reads which contain a region which maps to the genome, and either terminate in a string of A nucleotides or begin with a string of T nucleotides that do not map (and are thus characteristic of the poly-A tail), or by selecting for only these 3′ end fragments during either library preparation or sequencing [39, 40]. Differential alternative poly-A site usage can then be identified using standard *t*-test or Fisher's Exact tests. These direct approaches provide the strongest evidence for alternative polyadenylation; however, it can also be inferred by looking for a characteristic drop in read density at a specific point in the 3′UTR. Similar to analysis of alternative splicing events, programs such as MISO can use read density before and after a putative alternative polyadenylation site to estimate the percent usage of the two sites [38].

12 Transcriptome-wide Measurement of RNA Stability

Once an RNA transcript has been spliced, polyadenylated, and exported from the nucleus, the half-life of each mRNA molecule is tightly controlled through regulatory interactions that involve various RNA-binding proteins. These interactions

include targeted degradation mechanisms, such as microRNA-mediated RNA degradation through interactions with the Argonaute RNA-binding protein family and nonsense-mediated decay of transcripts containing premature stop codons that involves the activity of RBPs including UPF1, as well as more subtle regulation of RNA stability [41, 42]. AU-rich elements, one general class of RNA sequence element commonly found in 3′UTRs, can be bound by various factors that either stabilize the mRNA (including the HuD/HuR family of RBPs) or destabilize the RNA (including factors like AUF1) [42, 43].

Given an RNA-binding protein that regulates stability of a particular mRNA, the effect of knocking down that RBP will be to either increase (for destabilizing interactions) or decrease (for stabilizing interactions) the observed expression level of that mRNA. Thus, a simple gene expression quantification (either individually by qPCR, or done in high-throughput by RNA-seq or microarray) can serve as a first-pass analysis to identify the targets of an RBP regulated at the level of message stability. However, this analysis will also identify numerous differentially expressed genes that are secondarily regulated by altered transcription factors or other regulators and not true targets of the RBP itself. To address this limitation, specialized protocols like the BRIC-seq methodology have been developed to specifically quantitate RNA half-life (Fig. 1.1d) [44]. By metabolically labeling RNA with 5′-bromouridine (BrU) followed by a chase with fresh media, only RNA molecules transcribed until a specific timepoint are labeled. RNA is then sequenced at specified timepoints (0 h, 4 h, 8 h, and 12 h) post-chase to quantify the decrease in BrU incorporation level over time, which serves as a proxy for RNA half-life. Application of this method in HeLa cells identified significant differences in gene function among genes with long and short half-lives, suggesting that RNA stability may play critical roles in regulating specific cellular functions [44].

13 Global Quantification of Ribosome Occupancy

Recent technical advances have also enabled efforts to quantify altered regulation of translation. The ideal experiment, full proteome mass-spectrometry to quantify expression of all proteins, remains technically challenging and cost-prohibitive for most groups. However, methods to profile actively translating ribosomes have shown that ribosomal occupancy can serve as an effective proxy for translation rate, if not for protein expression itself (Fig. 1.1b) [45, 46]. Briefly, RNA is digested and 80S ribosomes (along with associated RNA fragments) are isolated by centrifugation through a sucrose cushion. Protected RNA fragments are then purified, ligated to oligonucleotide linkers, and amplified to generate libraries compatible with high-throughput sequencing [45]. This methodology is now available through the commercial ARTseq Ribosome Profiling Kit from Epicentre, making it possible for labs without previous expertise to characterize ribosomal occupancy genome-wide in a standardized manner. Further specialized protocols have made use of the biotin

ligase protein BirA to specifically tag ribosomes localized proximal to specific subcellular structures, allowing profiling of localized translation [47].

However, some specialized computational analysis is required to take full advantage of ribosome profiling data generated by high-throughput sequencing. At the core, analysis of these datasets faces similar challenges to CLIP-seq datasets; although profiling of translation is both experimentally as well as computationally more challenging than RNA profiling (by RNA-seq or microarray), it has two major advantages. First, numerous studies have suggested a surprisingly low correlation between mRNA and protein expression (whether raw levels or differential expression between two conditions) [48, 49], suggesting that profiling that is closer to protein expression itself will yield results that will correlate more closely with true biology. Second, recent studies have revealed that individual ribosomal subunits can show tissue-specific expression patterns and can play very specific roles in translation regulation. For example, large ribosomal subunit RPL38 appears to specifically regulates translation of HOX mRNAs, as mutation of RPL38 in mice has little effect on synthesis of other proteins but leads to significant developmental phenotypes characteristic of loss of HOX protein expression [50]. Thus, analysis that focuses solely on mRNA expression levels may miss significant translational regulation with significant roles in development and disease.

14 Challenges of Scale

Until recently, the approaches described above were largely used to analyze either single or small numbers of datasets. However, the combination of ENCODE-scale projects as well as the continual decrease in sequencing cost has enabled projects that profile RNA processing in hundreds, or even thousands, of conditions. Indeed, preparation of RNA-seq libraries in 96-well multiplexed format is now standardized in kit format from various commercial vendors. This increase in scale represents significant challenges; analysis pipelines that require on the order of 1 day per dataset do not represent a problem for small-scale experiments, but would be prohibitively long for sequential analysis of thousands of datasets. Thus, one of the major current computational challenges facing RNA analysis is to reformulate current approaches in ways that are faster and more rapidly scalable in order to handle large numbers of datasets.

Progress on this front is being made for many aspects of the various analysis pipelines. The availability of large computational clusters, and even commercial availability of multiple core desktop computers with 32–64 gb of RAM, have placed CPU time and storage space as the rate-limiting steps for primary dataset analyses. The STAR alignment program (developed as part of the ENCODE RNA-seq efforts) has made substantial progress in addressing read mapping, typically one of the slowest and most resource-intensive steps in RNA-seq analysis, increasing mapping speed by a factor of 50 above pre-existing methods at a cost of increased RAM usage that is still reasonably available in standard computer facilities [34].

Samtools, implementing the SAM and companion binary compressed BAM file formats, has enabled rapid post-mapping read statistics and alignment retrieval with a highly compressed file format that removes the need to store multiple large intermediate read alignment files [51]. Further work remains to optimize downstream analyses, such as CLIP-seq cluster identification, to increase the scalability of these resources as well.

The second major aspect of scaling is the automation of standard quality control and other basic analyses. With thousands of datasets, it is impractical to deeply explore each dataset to look for indications of contamination, low-complexity or over-amplified sequence libraries, or other potential biases. Thus, these large-scale analyses must include some level of automatic processing to flag potentially troublesome datasets for further analysis. In addition to basic technical quality control performed during high-throughput sequencing, programs such as FastQC incorporate a variety of calculations including per base and per read quality score, sequence bias, and over-represented k-mer analyses to flag potentially problematic datasets [52]. Efforts to optimize analysis-specific measures, such as the percent of CLIP-seq reads that fall into identified binding site clusters or the distribution of mismatched bases in iCLIP (that indicate the position of binding at single-base resolution) are ongoing.

Finally, efforts continue to develop methods of visualizing both quality control metrics as well as processed data from multiple datasets in parallel. FastQC provides useful tools to visualize the various quality control metrics, but developing ways to combine these reports into easy to view cross-sample reports remains an open challenge. RNA-SeQC, developed by the Cancer Genome Atlas, takes in highly compressed BAM files and provides a more comprehensive set of quality control metrics as both HTML and tab delimited files, helping rapid analysis of large numbers of datasets in parallel [53]. Similarly, visualizing large-scale post-analysis data in a useful manner remains challenging. Tools like the locally installed Integrated Genomics Viewer (IGV) [54] and web-based resources like the UCSC Genome Browser [55] provide simple frameworks in which multiple datasets can be viewed and compared. However, simultaneous viewing of many datasets (each of which could be gigabytes in size) is somewhat more challenging, depending on the ultimate goal. With sufficient computational resources, local installation of IGV (or similar browsers) is ideal for individual viewing, but lacks easy sharing across groups. In addition to the standard UCSC Genome Browser hosted by UCSC (which has limits on the number and size of datasets that can be uploaded and directly viewed), UCSC has also provided a mechanism to install the browser on a local server, providing a way to avoid data transfer issues by accessing private locally hosted datasets. Additionally, UCSC and others have recently developed a "Track Hub" system, by which groups can visualize a large number of genome-scale datasets in the normal UCSC browser framework while the raw data is hosted on a remote server. In this way, large-scale datasets can be viewed alongside other public data (including the ENCODE data that is publicly hosted at UCSC) while still maintaining data privacy and local storage.

15 Learning Predictive RNA Processing Networks

The identification of RBP binding sites (both *in vitro* and *in vivo*) can provide molecular insights into the basic mechanisms of RBP binding and function. However, it is infeasible to identify the targets of every RBP independently in all the various cell-types, tissues, and conditions in which the RBP may function. Thus, computational techniques need to be developed that can take target information generated in easy-to-manipulate cell lines and tissues, and integrate this information into predictive models of RBP regulation which can then be used to predict roles of an RBP in conditions not yet experimentally explored. Although challenging, research into these models has progressed along three tracks: first, generating predictive models of splicing by using sequence information alone; second, by deeply exploring the targets of an individual RBP, and third, by combining target information for multiple RBPs to infer combinatorial effects of multiple RBP regulators.

The ability to predict whether mutations or polymorphisms are likely to alter splicing, and ultimately isoform usage ratios themselves, would greatly aid rapid interpretation of potential disease-causal mutations identified in genome-wide association and whole-genome sequencing studies. Before it was experimentally possible to identify *in vivo* binding sites in a high-throughput and genome-wide manner, sequence motif enrichment and cross-species conservation provided the best opportunity to infer differential RNA regulation. By analyzing exonic and intronic sequences for motifs that are enriched nearby alternatively spliced exons, that showed unusual conservation across species, or that were enriched near splice sites, numerous efforts have identified a variety of short sequence motifs that correlate with splicing regulatory activity [29, 56]. Initially, these motifs (as well as other sequence features) were used to address a simple question: could alternatively spliced exons be distinguished from constitutively spliced exons based solely on sequence information? Using various machine learning models, various groups were able to perform such prediction with high accuracy in both mammals (human & mouse) as well as Drosophila, indicating that substantial regulatory information was contained within the RNA sequence itself [57–59].

Barash, *et al*. [60] extended this analysis to ask whether such sequence features could be used to not only predict whether an exon would be alternatively spliced or not, but also in which specific tissue such alternatively splicing would occur. Using a machine learning approach with 1014 different sequence features (including binding motifs for known factors as well as computationally identified motifs with yet unknown regulators, transcript structure features such as exon and intron size, etc.), Barash, *et al.* showed impressive accuracy in predicting tissue-specific alternative splicing patterns, including an ability to predict altered tissue-specific isoform usage upon mutation of putative regulatory elements [60]. Xiong *et al*. [61] extended this approach to the more basic question of predicting Psi values themselves by incorporating 1393 sequence features into a machine learning splicing predictor [61]. This method showed good accuracy for prediction of Psi values (R^2 of 0.65, with higher accuracy for predictions of increasing confidence), sufficient to perform the first

global analysis of the effect on splicing for all annotated single-nucleotide variants, identifying over 20,000 common and rare variants predicted to significantly alter splicing regulation. The results of this and other efforts suggest that an ability to rapidly and accurately predict the effect of novel mutations on splicing will be achieved in the near future [60–62].

16 Integrating Target Information to Generate Regulatory Maps for Individual RBPs

Second, detailed analyses of the targets of individual RNA-binding proteins can give an in-depth view of the differential regulatory activities an RBP can have in different contexts. Perhaps the most well-studied example of detailed characterization of the differential effect of RBP interaction at various positions in exons and flanking introns is the activity of neuronal splicing factors NOVA1 and NOVA2. NOVA proteins show specific expression patterns in the brain, and are essential for postnatal motor neuron survival [63, 64]. Various early molecular studies, including *in vitro* binding assays and X-ray crystallography, indicated that the KH-type RNA-binding domains of NOVA directly interact with YCAY sequence motif clusters [64–66]. Although alternative splicing of many exons were known to be regulated by NOVA, NOVA induced inclusion of some but exclusion of others, suggesting that a simple presence of YCAY clusters was not sufficient to predict the mode of NOVA regulation.

To better understand NOVA-mediated splicing regulation, Ule. *et al.* [67] developed the concept of a 'splicing map', in which they calculated the frequency of YCAY clusters at various positions within NOVA-dependent exons as well as their flanking introns [67]. The degree of enrichment was then plotted on a generic, size-normalized exon-intron-exon structure to create two separate motif-enrichment plots: one for NOVA-dependent included exons, and another for NOVA-dependent excluded exons. These maps showed striking patterns of differential regulation by NOVA based on YCAY location; YCAY clusters within the alternative exon or located near the 5′ splice site of the upstream intron correlated with silencing activity, whereas YCAY clusters near the 5′ or 3′ splice sites in the downstream exon correlated with enhancing activity. Although this analysis was performed using sequence information alone, later work using *in vivo* binding sites for NOVA identified by a CLIP-seq-like methodology yielded a nearly identical splicing map for exons directly bound by NOVA [12]. Importantly, these maps were able to predict novel NOVA-dependent regulated exons, with 30 of 51 predicted alternative exons showing NOVA-dependent regulation (all 30 of which were correctly predicted as NOVA-enhancing or NOVA-repressing respectively) [67].

Following this work, other analyses created such splicing maps for a variety of factors that revealed further insight into RNA regulatory processes. Some factors, such as RBFOX2, show similar characteristics to the maps generated for NOVA:

RBFOX2 binds to a specific UGCAUG element which is associated with exon silencing when present in 3′ end of the upstream intron, but associated with enhanced exon inclusion if located near the 5′ splice site [13]. In contrast, analysis of HNRNPC *in vivo* binding by iCLIP indicated enrichment near both the 5′ and 3′ ends of hnRNP C-silenced exons. In this case, further analysis focusing on only those loci with 160–170 nt of intervening sequence between two cross-linked nucleotides (which is indicative of full hnRNP particle formation) revealed further subtlety; full hnRNPC particles were associated with exon inclusion in the upstream intron, but with silencing when bound to the exon itself [9]. RNA splicing maps have now been generated for a variety of RNA-binding proteins, revealing interesting similarities and differences in their effect at various positions along an alternative exon and its flanking introns [68].

However, this simple model does not suffice for all factors. Analysis of *in vivo* targets of ALS-linked RNA-binding protein FUS/TLS (fused in sarcoma/translocated in liposarcoma) revealed a unexpectedly broad pattern of binding. FUS binding was highly enriched for binding within long introns, where it exhibited a characteristic 'sawtooth-like' pattern (with highest association observed at the 5′ splice site but decreasing throughout the intron) suggestive of co-transcriptional deposition. For FUS, intronic length proved to be one of the key features of whether a bound target would be differentially expressed upon FUS knockdown, suggesting that functional splicing maps for some RBPs may require a more complex model than simply localization of binding within the intron-alternative exon-intron cassette [69].

The generation of RNA splicing maps has provided a useful analysis method to understand the roles of splicing regulators. The same type of analysis, incorporating either RNA half-life or ribosome occupancy, can similarly provide insight into the location-dependent roles an RBP plays in regulating other aspects of RNA processing. For example, incorporation of CLIP-seq, ribosome footprinting, and RNA-seq data for NMD factor UPF1 revealed that in addition to the traditional features characteristic of RNAs targeted for degradation by NMD (long 3′ UTRs, presence of a premature stop codon upstream of the terminal exon junction, and presence of an upstream open reading frame), identification of upstream open reading frames with significant ribosome footprinting signal provided improved ability to predict NMD targets [70]. Additionally, UPF1 binding was predictive of degradation regardless of 3′UTR length, suggesting that the correlation between 3′UTR length and NMD may be either upstream or independent of UPF1-mediated degradation [70]. Further analysis using both RIP- and CLIP-seq of UPF1 ATPase-mutants revealed that it was not UPF1 recruitment to specific NMD targets, but rather regulation of UPF1 dissociation from non-target mRNAs, that controls which mRNAs will be ultimately degraded [71]. Although profiling of translation is not yet as commonly performed as RNA quantification, the recent commercial availability of ribosome footprinting protocols should enable further efforts to link RBP regulatory activity to not only RNA expression, but translational levels as well.

17 Integration of Multiple RBP Datasets

Integrated analysis of multiple RNA-binding proteins in parallel presents an opportunity to learn the regulatory networks for individual proteins, the interplay between co-regulating factors, and build more complete RNA regulatory networks to explain altered RNA processing in a cell-type or disease of interest. Many efforts along these lines have focused on coordinated studies of paralogous factors, which often show complex co-regulation of targets.

The degree of functional redundancy can vary greatly among RNA-binding protein families. Identification of *in vivo* targets of the three FMR1 RNA-binding protein family members (FMRP, FXR1, and FXR2) revealed that >95 % of FXR1 and FXR2 binding sites co-localized with FMRP binding sites [72]. Similarly, analysis of MBNL1 and MBNL2 targets in various tissues across human and mouse suggested redundant roles in splicing regulation [73, 74]. In contrast, target identification of other factors revealed striking differences between binding patterns. RNA-binding proteins TDP-43 and FUS/TLS have both been shown to play independent significant roles in amyotrphic lateral sclerosis (ALS), as mutations in both have been proposed to cause a significant fraction of inherited ALS cases [75]. Although yielding similar phenotypes, identification of *in vivo* targets (by CLIP-seq) as well as identification of differentially regulated targets (identified by knockdown of TDP-43 and FUS/TLS in *in vitro* differentiated human neurons) showed largely distinct sets of targets at both the gene expression and splicing level [69]. However, by focusing on the small number of genes that were bound by both TDP-43 and FUS/TLS, it was found that a small number of transcripts encoding genes with essential functions in neurons were downregulated upon knockdown of either TDP-43 or FUS/TLS, suggesting potential co-regulation of a small but biologically relevant subset of targets [69].

Intriguingly, other work has shown that binding interactions themselves can be dependent upon the combinatorial effect of other RNA-binding proteins. Analysis of SRSF1 binding sites in cells in which SRSF2 is depleted by siRNA suggested that SRSF1 and SRSF2 compete for binding at similar sites, as SRSF1 binding was enriched at strong SRSF2 binding sites upon SRSF2 depletion [76]. However, SRSF2 depletion led to reduction of SRSF1 binding at other sites, suggesting that at other (typically weaker) binding sites SRSF1 and SRSF2 binding is coupled through an unknown mechanism [76].

The emergence of high-throughput target identification methods enables large-scale cross-protein comparisons. The first such effort in *S. cerevisiae* profiled *in vivo* targets for 40 RNA-binding proteins, including a number of proteins not previously annotated as interacting with RNA. The generation of these datasets using identical cells and methodology enabled cross-protein analyses, identifying novel co-binding between proteins and the first semi-global look at the combinatorial binding of dozens of RNA-binding proteins in parallel [77]. Similar efforts to profile RNA-binding protein targets for many proteins in parallel have identified targets of four hnRNP proteins (hrp36, hrp38, hrp40, and hrp48) in *Drosophila*, 12 assorted RNA-binding

proteins (PUM2, QKI, IGF2BP1-3, AGO1-4, and TNRC6A-C) in human HEK293 cells, six HNRNP family members (HNRNPA1, HNRNPA2/B1, HNRNPF, HNRNPH1, HNRNPM, and HNRNPU) in human 293 T cells, and four eIF3 translation initiation complex members (eIF3a, eIF3b, eIF3d, and eif3G) in human 293 T cells [10, 30, 78, 79]. In each case, the generation of target information in a standardized manner in the same cell-type enabled cross-dataset comparisons that revealed unexpected complexity in factor co-association and co-regulation of targets. These types of analyses represent both opportunities and challenges: the opportunity to obtain the first truly global views of RNA processing regulation, but the challenge of developing computational methods to integrate not only multiple experiments for an individual RBP, but to also analyze dozens (or hundreds) of RBPs in parallel. This includes the significant complexity inherent in analyzing RNA-binding proteins that are often alternatively spliced or otherwise regulated at the RNA level themselves, leading to substantial cross-regulation among RBPs [30]. Although these initial analyses have largely focused on individual RBP regulation with a brief consideration for combinatorial regulation, the next stage of these efforts will likely turn to machine learning algorithms to help to fully understand the fully complexity of the human RNA processing regulatory network. The development of these methods will require large resources of RBP targets, such as the over one hundred RBPs profiled by eCLIP in K562 and HepG2 human cell lines as part of the ongoing ENCODE consortium efforts [15], to properly train and validate such approaches.

In addition to providing RBP-specific insights, integrative analysis may reveal previously unknown properties of RNA processing and regulation. An interesting parallel can be seen for studies of transcription factor targets: although identification of *in vivo* binding sites for many individual transcription factors (by chromatin immunoprecipitation followed by high-throughput sequencing, or "ChIP-seq") has yielded significant insights into the biological roles of those individual factors, large-scale ChIP-seq efforts performed by individual labs as well as the ENCODE project (in human) and modENCODE project (in *Drosophila* and *C. elegans*) revealed novel general properties of transcriptional regulation [80, 81]. In particular, these analyses led to the characterization of HOT regions (loci bound by the majority of assayed factors) as a novel regulatory mechanism for regulation of essential housekeeping genes [80–82], and led to the development of models to predict target gene expression based on a variety of features (including both sequence and binding information) [83]. These results strongly suggest that in addition to RBP-specific insights, it remains possible that these large-scale efforts to profile RNA regulatory networks may also reveal completely novel principles in regulation of RNA processing.

18 Conclusion

Over the coming years, our knowledge of RBP targets will continue to rapidly expand. In addition to the efforts of individual labs, which have identified targets for dozens of RBPs in various cell lines and tissues and will continue to deeply explore RNA processing regulation, the ENCODE project has now added an effort to profile

the targets of hundreds of annotated or predicted RNA-binding proteins in two standard laboratory human cell lines. With these large-scale data generation efforts come significant computational challenges in automatically processing these datasets, subjecting them to automated quality control procedures, analyzing thousands of datasets in a rapid yet accurate manner, visualizing these analyses, and (most importantly) incorporating this information in order to infer novel aspects of RNA biology. Although many challenges remain, the work of numerous groups over the past few years have revealed significant insights into how RNA-binding proteins act to regulate RNA processing across various tissues and cell-types, and how alteration of these regulatory activities can lead to disease. The further development of computational tools in the coming years will allow a rapid expansion of this research, enabling a global picture of RNA processing that should lead to further insights into the roles that RNA processing plays in defining gene expression in humans.

Acknowledgements This work was partially supported by grants from the National Institute of Health (HG007005, HG004659 and NS075449) to G.W.Y. G.W.Y. is an Alfred P. Sloan Research Fellow. E.L.V.N. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2172-13).

References

1. Muller-McNicoll M, Neugebauer KM (2013) How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat Rev Genet 14:275–87. doi:10.1038/nrg3434
2. 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. doi:10.1371/journal.pone.0015499
3. Tsvetanova NG, Klass DM, Salzman J, Brown PO (2010) Proteome-wide search reveals unexpected RNA-binding proteins in Saccharomyces cerevisiae. PLoS One 5. doi:10.1371/journal.pone.0012671
4. Castello A et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–406. doi:10.1016/j.cell.2012.04.031
5. Kwon SC et al (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol 20(9):1122–30. doi:10.1038/nsmb.2638
6. Paige JS, Wu KY, Jaffrey SR (2011) RNA mimics of green fluorescent protein. Science 333:642–6. doi:10.1126/science.1207339
7. Nicol C et al (2013) An RNA aptamer provides a novel approach for the induction of apoptosis by targeting the HPV16 E7 oncoprotein. PLoS One 8, e64781. doi:10.1371/journal.pone.0064781
8. Shui B et al (2012) RNA aptamers that functionally interact with green fluorescent protein and its derivatives. Nucleic Acids Res 40, e39. doi:10.1093/nar/gkr1264
9. Konig J et al (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–15. doi:10.1038/nsmb.1838
10. Hafner M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–41. doi:10.1016/j.cell.2010.03.009
11. Ule J et al (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–5. doi:10.1126/science.1090095
12. Licatalosi DD et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–9. doi:10.1038/nature07488
13. Yeo GW et al (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol 16:130–7. doi:10.1038/nsmb.1545

14. Zhang C, Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29:607–14. doi:10.1038/nbt.1873
15. Van Nostrand EL et al (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods 2016 Mar 28 advance online publication. doi: 10.1038/nmeth.3810
16. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497–502. doi:10.1126/science.1141319
17. Valouev A et al (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5:829–34. doi:10.1038/nmeth.1246
18. Cooper TA (2005) Use of minigene systems to dissect alternative splicing elements. Methods 37:331–40. doi:10.1016/j.ymeth.2005.07.015
19. Lambeth LS, Smith CA (2013) Short hairpin RNA-mediated gene silencing. Methods Mol Biol 942:205–32. doi:10.1007/978-1-62703-119-6_12
20. Rao DD, Senzer N, Cleary MA, Nemunaitis J (2009) Comparative assessment of siRNA and shRNA off target effects: what is slowing clinical development. Cancer Gene Ther 16:807–9. doi:10.1038/cgt.2009.53
21. Ho TH et al (2004) Muscleblind proteins regulate alternative splicing. EMBO J 23:3103–12. doi:10.1038/sj.emboj.7600300
22. Tang ZZ, Zheng S, Nikolic J, Black DL (2009) Developmental control of CaV1.2 L-type calcium channel splicing by Fox proteins. Mol Cell Biol 29:4757–65. doi:10.1128/MCB.00608-09
23. Gehman LT et al (2011) The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet 43:706–11. doi:10.1038/ng.841
24. McPherson LA, Weigel RJ (1999) AP2alpha and AP2gamma: a comparison of binding site specificity and trans-activation of the estrogen receptor promoter and single site promoter constructs. Nucleic Acids Res 27:4040–9
25. Sarkar B, Lu JY, Schneider RJ (2003) Nuclear import and export functions in the different isoforms of the AUF1/heterogeneous nuclear ribonucleoprotein protein family. J Biol Chem 278:20700–7. doi:10.1074/jbc.M301176200
26. Johnson JM et al (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science 302:2141–4. doi:10.1126/science.1090100
27. Shen S, Warzecha CC, Carstens RP, Xing Y (2010) MADS+: discovery of differential splicing events from Affymetrix exon junction array data. Bioinformatics 26:268–9. doi:10.1093/bioinformatics/btp643
28. Shai O, Morris QD, Blencowe BJ, Frey BJ (2006) Inferring global levels of alternative splicing isoforms using a generative model of microarray data. Bioinformatics 22:606–13. doi:10.1093/bioinformatics/btk028
29. Yeo GW, Van Nostrand EL, Liang TY (2007) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. PLoS Genet 3, e85. doi:10.1371/journal.pgen.0030085
30. Huelga SC et al (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1:167–78. doi:10.1016/j.celrep.2012.02.001
31. Du H et al (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nat Struct Mol Biol 17:187–93. doi:10.1038/nsmb.1720
32. Chen L (2011) Statistical and computational studies on alternative splicing. In: Schölkopf B, Lu HH-S, Zhao H (eds) Handbook of statistical bioinformatics. Springer, Berlin, pp 31–53
33. Parkhomchuk D et al (2009) Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37, e123. doi:10.1093/nar/gkp596
34. Dobin A et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi:10.1093/bioinformatics/bts635
35. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi:10.1186/gb-2009-10-3-r25
36. Trapnell C et al (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31:46–53. doi:10.1038/nbt.2450

37. Wang ET et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–6. doi:10.1038/nature07509
38. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 7:1009–15. doi:10.1038/nmeth.1528
39. Shepard PJ et al (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17:761–72. doi:10.1261/rna.2581711
40. Derti A et al (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res 22:1173–83. doi:10.1101/gr.132563.111
41. Hutvagner G, Simard MJ (2008) Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9:22–32. doi:10.1038/nrm2321
42. Houseley J, Tollervey D (2009) The many pathways of RNA degradation. Cell 136:763–76. doi:10.1016/j.cell.2009.01.019
43. Lal A et al (2004) Concurrent versus individual binding of HuR and AUF1 to common labile target mRNAs. EMBO J 23:3092–102. doi:10.1038/sj.emboj.7600305
44. Tani H et al (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res 22:947–56. doi:10.1101/gr.130559.111
45. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–23. doi:10.1126/science.1168978
46. 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. doi:10.1016/j.cell.2011.10.002
47. Jan CH, Williams CC, Weissman JS (2014) Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science 346:1257521-1-8. doi: 10.1126/science.1257521
48. Schwanhausser B et al (2011) Global quantification of mammalian gene expression control. Nature 473:337–42. doi:10.1038/nature10098
49. Gry M et al (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. BMC Genomics 10:365. doi:10.1186/1471-2164-10-365
50. Kondrashov N et al (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 145:383–97. doi:10.1016/j.cell.2011.03.028
51. Li H et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–9. doi:10.1093/bioinformatics/btp352
52. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc
53. DeLuca DS et al (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28:1530–2. doi:10.1093/bioinformatics/bts196
54. Robinson JT et al (2011) Integrative genomics viewer. Nat Biotechnol 29:24–6. doi:10.1038/nbt.1754
55. Kent WJ et al (2002) The human genome browser at UCSC. Genome Res 12:996–1006. doi:10.1101/gr.229102
56. Fairbrother WG, Yeh RF, Sharp PA, Burge CB (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297:1007–13. doi:10.1126/science.1073774
57. Dror G, Sorek R, Shamir R (2005) Accurate identification of alternatively spliced exons using support vector machine. Bioinformatics 21:897–901. doi:10.1093/bioinformatics/bti132
58. Sorek R et al (2004) A non-EST-based method for exon-skipping prediction. Genome Res 14:1617–23. doi:10.1101/gr.2572604
59. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB (2005) Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A 102:2850–5. doi:10.1073/pnas.0409742102
60. Barash Y et al (2010) Deciphering the splicing code. Nature 465:53–9. doi:10.1038/nature09000
61. Xiong HY et al (2015) RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science 347:1254806. doi:10.1126/science.1254806

62. Faber K, Glatting KH, Mueller PJ, Risch A, Hotz-Wagenblatt A (2011) Genome-wide prediction of splice-modifying SNPs in human genes using a new analysis pipeline called AASsites. BMC Bioinformat 12(Suppl 4):S2. doi:10.1186/1471-2105-12-S4-S2
63. Yang YY, Yin GL, Darnell RB (1998) The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. Proc Natl Acad Sci U S A 95:13254–9
64. Jensen KB et al (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron 25:359–71
65. Buckanovich RJ, Darnell RB (1997) The neuronal RNA-binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. Mol Cell Biol 17:3194–201
66. Lewis HA et al (2000) Sequence-specific RNA-binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. Cell 100:323–32
67. Ule J et al (2006) An RNA map predicting Nova-dependent splicing regulation. Nature 444:580–6. doi:10.1038/nature05304
68. Witten JT, Ule J (2011) Understanding splicing regulation through RNA splicing maps. Trends Genet 27:89–97. doi:10.1016/j.tig.2010.12.001
69. Lagier-Tourenne C et al (2012) Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat Neurosci 15:1488–97. doi:10.1038/nn.3230
70. Hurt JA, Robertson AD, Burge CB (2013) Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. Genome Res 23:1636–50. doi:10.1101/gr.157354.113
71. Lee SR, Pratt GA, Martinez FJ, Yeo GW, Lykke-Andersen J (2015) Target discrimination in nonsense-mediated mRNA decay requires Upf1 ATPase activity. Mol Cell 59:413–25. doi:10.1016/j.molcel.2015.06.036
72. Ascano M Jr et al (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature 492:382–6. doi:10.1038/nature11737
73. Wang ET et al (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. Cell 150:710–24. doi:10.1016/j.cell.2012.06.041
74. Han H et al (2013) MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature 498:241–5. doi:10.1038/nature12270
75. Lagier-Tourenne C, Polymenidou M, Cleveland DW (2010) TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum Mol Genet 19:R46–64. doi:10.1093/hmg/ddq137
76. Pandit S et al (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol Cell 50:223–35. doi:10.1016/j.molcel.2013.03.001
77. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6, e255. doi:10.1371/journal.pbio.0060255
78. Blanchette M et al (2009) Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell 33:438–49. doi:10.1016/j.molcel.2009.01.022
79. Lee AS, Kranzusch PJ, Cate JH (2015) eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. Nature 522:111–4. doi:10.1038/nature14267
80. Gerstein MB et al (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330:1775–87. doi:10.1126/science.1196914
81. modENCODE Consortium et al (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330:1787–97. doi:10.1126/science.1198374
82. Gerstein MB et al (2012) Architecture of the human regulatory network derived from ENCODE data. Nature 489:91–100. doi:10.1038/nature11245
83. Marbach D et al (2012) Predictive regulatory models in Drosophila melanogaster by integrative inference of transcriptional networks. Genome Res 22:1334–49. doi:10.1101/gr.127191.111

Chapter 2
Genome-Wide Approaches for RNA Structure Probing

Ian M. Silverman, Nathan D. Berkowitz, Sager J. Gosai, and Brian D. Gregory

Abstract RNA molecules of all types fold into complex secondary and tertiary structures that are important for their function and regulation. Structural and catalytic RNAs such as ribosomal RNA (rRNA) and transfer RNA (tRNA) are central players in protein synthesis, and only function through their proper folding into intricate three-dimensional structures. Studies of messenger RNA (mRNA) regulation have also revealed that structural elements embedded within these RNA species are important for the proper regulation of their total level in the transcriptome. More recently, the discovery of microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) has shed light on the importance of RNA structure to genome, transcriptome, and proteome regulation. Due to the relatively small number, high conservation, and importance of structural and catalytic RNAs to all life, much early work in RNA structure analysis mapped out a detailed view of these molecules. Computational and physical methods were used in concert with enzymatic and chemical structure probing to create high-resolution models of these fundamental biological molecules. However, the recent expansion in our knowledge of the

I.M. Silverman
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

N.D. Berkowitz
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

S.J. Gosai
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

B.D. Gregory (✉)
Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA
e-mail: bdgregor@sas.upenn.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_2

importance of RNA structure to coding and regulatory RNAs has left the field in need of faster and scalable methods for high-throughput structural analysis. To address this, nuclease and chemical RNA structure probing methodologies have been adapted for genome-wide analysis. These methods have been deployed to globally characterize thousands of RNA structures in a single experiment. Here, we review these experimental methodologies for high-throughput RNA structure determination and discuss the insights gained from each approach.

Keywords PARS • FragSeq • ds/ssRNA-seq • DMS-seq • Structure-seq • CIRS-seq • MOD-seq • hSHAPE • SHAPE-CE • SHAPE-seq

1 Introduction

Among the biological macromolecules, RNA has the distinction of both being both a carrier of genetic information (like DNA) as well as a catalytic machine (like protein). The primary sequence information of RNA carries the genetic code from DNA to the ribosome to direct protein synthesis. The sequence of RNA also underlies its secondary structure, the pattern of hydrogen bonds that connect non-adjacent bases. Such bonds can be formed through Watson-Crick and Hoogsteen base-pair interactions. These interactions, in concert with auxiliary factors and coordinating ions, enable RNAs to form complex tertiary structures that carry out diverse enzymatic and regulatory processes [1].

Indeed, processes fundamental to life, including translation, are dependent on specific and complex RNA structures. For instance, transfer RNAs (tRNAs) adhere to a distinctive secondary structure reminiscent of a cloverleaf, which is nearly ubiquitously conserved [2]. This mediates folding of these molecules into an L-shaped tertiary structure so they can funnel into the ribosome active site during translation. The ribosome itself is also a complex molecular machine made up of numerous RNAs and proteins, and its shape and function is entirely dependent on the precise folding of its RNA components [3].

RNA structural moieties have also evolved within the context of other functional RNA molecules. For instance, riboswitches are small structural elements that change conformation in response to direct binding to a specific metabolite, which ultimately affects RNA stability or translation efficiency [4]. Dozens of riboswitches have been discovered in prokaryotes, but thus far only the thiamine pyrophosphate (TPP) riboswitch has been found in eukaryotes [5]. This raises the question of why riboswitches, one of the most ancient and conserved gene regulatory mechanisms, have been so elusive in eukaryotes? This is likely because of the small number and more distant relationships of available sequenced eukaryotic genomes, which limits the statistical power needed to identify structural elements using a sequence-directed search.

While riboswitches themselves haven't been found in eukaryotes, small regulatory structural elements embedded within mRNAs have been identified that interact

with non-metabolite factors. For instance, the Iron Responsive Element (IRE), is found in numerous eukaryotic mRNAs encoding iron metabolism genes and interacts with IRE-binding protein in an iron-dependent manner to regulate gene expression [6]. Histone stem loops serve as stability factors for non-polyadenylated histone mRNAs in eukaryotes by recruitment of cofactors [7]. Finally, the internal ribosome entry site (IRES) is a structural element that allows for cap-independent translation, and is used by viral RNAs to hijack the infected cells translation machinery [8].

RNA structure is also central to the biogenesis and function of two recently discovered RNA classes, microRNAs (miRNAs) and long noncoding RNAs (lncRNAs). miRNAs, which regulate gene expression by inhibiting translation and promoting degradation of target mRNAs, are processed from long hairpin structures found in precursor transcripts. lncRNAs show poor sequence conservation but impart a strong regulatory function, which is mediated through secondary structures and/or interaction with trans-acting factors, such as RNA-binding proteins (RBPs) and chromatin [9]. Indeed, a detailed analysis of the steroid receptor RNA activator (SRA) structure demonstrated that lncRNAs have structural domains that are maintained through evolutionary pressures in lieu of primary sequence conservation [10].

As outlined above, the secondary structure of RNA is critical to its function and in turn to all biological life. Therefore, there has been widespread interest in developing methods to study RNA structures, in an effort to gain a better understanding of the mechanisms by which it functions. These methodologies fall into three main categories: physical, computational, and biochemical/molecular biological approaches.

In the 1970s the first complete RNA structures were solved using X-ray crystallography [11, 12]. This technique can produce extremely high quality structures and is often considered to be the gold standard, however there are technical and analytical caveats. Mainly, it requires that the RNA form highly ordered crystals, which is challenging for some RNAs and may be impossible for others, especially RNAs for which the structure is heterogeneous or dynamic [13]. Nuclear magnetic resonance (NMR), another physical method, is able to probe RNA structure in solution and can capture the dynamics of this feature [14]. However, like X-ray crystallography, it is only able to study RNAs in an artificial (*in vitro*) context. In total, both of these techniques are powerful for probing structure, but the main drawback they share is that their scale is limited.

Purely computational methods can be very rapid and require only the primary sequence of the RNA, which in the post-genomic world is readily available. Such approaches use thermodynamic models and try to minimize free energy values to model the most probable structure of RNAs [15, 16]. Computational tools have also been developed that can utilize sequence conservation to prioritize bases that have been maintained through evolution to retain specific structural confirmations [17]. However, *in silico*, sequence-based approaches can require large amounts of computing power, and it's not always clear that they take into account all relevant assumptions for a given RNA [18].

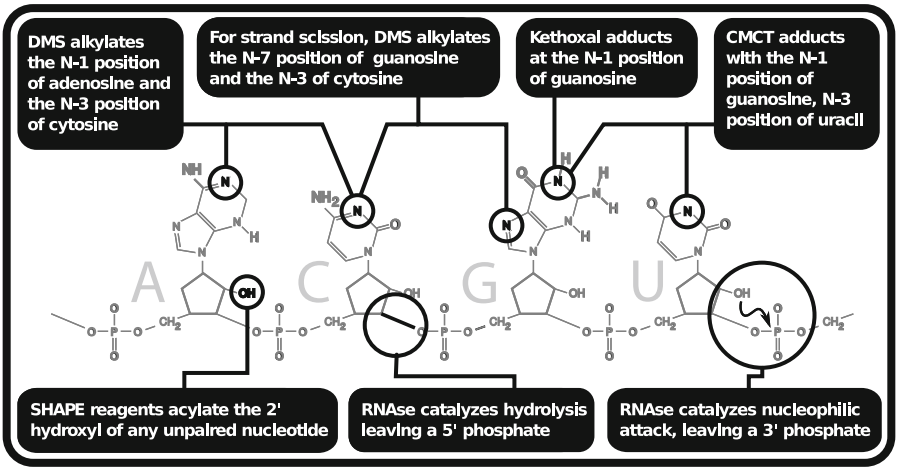


Fig. 2.1 Structure sensitive probes target diverse chemical groups. DMS, Kethoxal and CMCT adduct to various nitrogen atoms on the nucleoside bases. SHAPE reagents acylate the ribose backbone. RNases catalyze cleavage of the phosphodiester linkage between nucleotides

The third set of methods, which we survey in depth here, involves manipulating the chemical properties of RNA with an enzyme or small molecule. A wide variety of enzymatic and chemical probes can provide experimentally based structure information, and can in theory be used to study any transcript or transcriptome (Fig. 2.1 and Table 2.1). Their unifying property is that each reacts differentially with paired and unpaired regions of RNA. However, traditional implementations of this strategy have been limited by throughput and *in vivo* RNA abundance, especially when probing native RNA conformation. Although each technique uses a different set of reagents and has unique caveats, they all have the potential to reveal structural information for thousands of molecules in a single experiment when coupled to high-throughput sequencing (HTS) (Fig. 2.2). The focus of this chapter is specifically on experimental enzymatic and chemical methods for probing RNA structure and their use in combination with HTS technologies to give global views of RNA secondary structure in transcriptomes.

2 Classical RNA Structure Probing Methodologies

2.1 Nuclease-Based Approaches

Nearly five decades ago, the first realization that nucleases had structural specificity was born out of studies of tRNAs, which concluded that RNases purified from pancreatic tissue cleaved specifically at tRNA anticodons [19]. Since then,

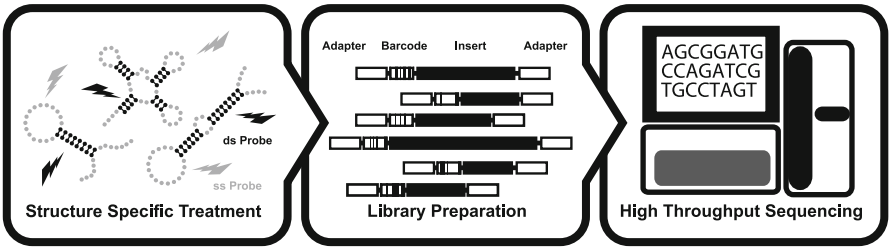


Fig. 2.2 General strategy for genome-wide RNA structure probing. A starting RNA population is treated with single-stranded or double-stranded probing reagent, followed by strand-specific RNA library preparation and high-throughput sequencing

enzymatic approaches have been widely used to probe RNA secondary structure in solution. These methods all rely on the same general principle, whereby a RNA of interest is subjected to hydrolysis by a structure-specific endonuclease which creates a cleavage pattern indicative of an RNAs structure (Fig. 2.1). To readout the digestion patterns, RNAs are first labeled at the 5′ end with γ-P^{32}-ATP by the action of T4 polynucleotide kinase (PNK), subjected to RNase treatment, and then directly analyzed by polyacrylamide-urea gel electrophoresis and autoradiography. Alternatively, primer extension with a radiolabeled primer can detect cleavage sites as reverse transcriptase (RT) stop sites followed by Sanger sequencing analysis [20]. In the latter approach, proper controls must be performed, due to the fact that naturally modified nucleotides and strong RNA structures can also block the elongation of RT, resulting in termination sites that will result in false positive structure calls.

There are a number of nucleases that have been used for RNA secondary structure analysis (Table 2.1). They vary in their structure- and sequence-specificity, size, optimal pH, mechanism of action, and requirement for cations [20]. RNase A, T1, and U2 are among the most commonly used nucleases in RNA structure probing experiments. All three enzymes cleave single-stranded regions via nucleoside 2′,3′-cyclic monophosphate intermediates, resulting in a 5′ OH and 3′ P (Table 2.1 and Fig. 2.1) [21–23]. However, each protein has specific nucleotide preferences for cleavage site selection. For instance, RNase A cleaves 3′ of pyrimidines, RNase T1 specifically cleaves 3′ of guanosine residues, and RNase U2 cleaves 3′ of adenines or guanosines (Table 2.1). Therefore, a comprehensive RNA structure probing experiment with these enzymes requires the use of multiple RNases to thoroughly analyze RNA structures.

There are only a handful of ribonucleases that show structural preference and no sequence preference. S1 Nuclease, Mung Bean (MB) nuclease, P1 nuclease, and RNase I are among the few nucleases with specific activity for single-stranded regions and no preference for particular sequences, making them ideal for RNA structure mapping [24, 25]. S1 and P1 nuclease cleavage results in a product with a 5′ P and 3′ OH and are reported to have optimal activity in acidic pH [25, 26]. MB

Table 2.1 Overview of classical approaches for RNA structure probing

Paradigm	Probe	RSS specificity	Nucleotide bias	Product(s)	Readout	Cell permeable
Nuclease	RNase A	ssRNA	Cp or Up	5′ OH and 3′ P	End labeling or primer extension	No
	RNase T 1	ssRNA	Gp	5′ OH and 3′ P	End labeling or primer extension	No
	RNase U2	ssRNA	Ap or Gp	5′ OH and 3′ P	End labeling or primer extension	No
	Mb Nuclease	ssRNA	None	5′ P and 3′ OH	End labeling or primer extension	No
	Nuclease P1	ssRNA	None	5′ P and 3′ OH	End labeling or primer extension	No
	Nuclease S1	ssRNA	None	5′ P and 3′ OH	End labeling or primer extension	No
	RNase I	ssRNA	None	5′ OH and 3′ P	End labeling or primer extension	No
	RNase V1	dsRNA	None	5′ P and 3′ OH	End labeling or primer extension	No
Base modification	DMS (RT)	ssRNA	A,C	N-Alkylation	Primer extention	Yes
	DMS (SS)	ssRNA	C,G	N-Alkylation	Strand scission and end labeling	Yes
	CMCT	ssRNA	G,U,Ψ	N-1 or N-3 Adduct	Primer extention	No (Yes if Permeabilized)
	kethoxal	ssRNA	G	N-1 Adduct	Primer extention	No (Yes if Permeabilized)
Backbone mod	1M7	ssRNA	None	2′ OH Acylation	Primer extension	No
	NMIA	ssRNA	None	2′ OH Acylation	Primer extension	No
	BzCN	ssRNA	None	2′ OH Acylation	Primer extension	No

nuclease also leaves 5′ P and 3′ OH and is optimally active at physiological pH. RNase I is highly active at physiological pH, however it cleaves RNA using a distinct mechanism, via nucleoside 2′,3′-cyclic monophosphate intermediates, resulting in a 5′ OH and 3′ P (Table 2.1).

In 1981, two groups independently reported the isolation and use of a nuclease isolated from the venom of *Naja oxiana* (Caspian Cobra), with specificity for double-stranded RNA (dsRNA) and no primary sequence preferences [27, 28]. RNase V1, as it is now known, cleaves 4–6 nucleotide stretches of structured RNAs to leave 5′ P and 3′ OH [29]. RNase V1 cleaves specifically double-stranded regions and has no sequence preference, and therefore has become a powerful reagent in the battery of enzymes used for RNA structure mapping (Table 2.1). RNase III type ribonucleases are also capable of cleaving dsRNA without sequence preference, however these enzymes require long stretches of dsRNA, such as rRNA or pre-miRNAs, making them not applicable to structural studies [30].

When multiple nucleases, with distinct structural and sequence preferences are used together and in concert with chemical probes (discussed below), these reagents provide a powerful collection of tools for structural analysis of almost any RNA. However, there are several caveats associated with nuclease-based methods. Primarily, the large size of enzymes compared to small molecules, means that steric hindrance could pose a problem for nucleases, particularly near strong structural elements. Additionally, structure-specific nucleases may not cleave all paired or unpaired sites equally due to the structure of the surrounding nucleotides. Finally, due to their size, these enzymes cannot permeate directly through the cell membrane.

2.2 Base Modification-Based Approaches

For decades, small molecules have been used to experimentally probe RNA secondary structure. For instance, agents that modify solvent accessible nucleobases at the atoms involved in Watson-Crick hydrogen bonding have been particularly useful (Fig. 2.1). In initial studies, RNA end-labeled with a radioactive phosphate was treated with dimethyl sulfate (DMS) to alkylate unpaired guanosines and cytidines or treated with diethyl pyrocarbonate to carbethoxylate unpaired adenosines [31, 32]. Strand scission was then induced at the modified base using aniline. Fragment lengths were identified by PAGE-autoradiography, and were used to identify modification sites that were thus determined to be unpaired.

While useful, this method was restricted to short stretches of RNA close to the radioactive RNA label. Subsequently, these modifications were found to terminate cDNA elongation by reverse transcriptase (RT) at the nucleotide preceding the modified base [33, 34]. Therefore, structure could be probed along the entirety of an RNA molecule by primer extension and electrophoretic fractionation.

DMS and 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide (CMCT), have been used extensively to probe RNA secondary structure using this strategy [20, 33–41]. More specifically, DMS treatment alkylates the N1 and N3 positions of adenine and cytosine, respectively [42], while CMCT results in adduct addition at the N3 position of uracil, N1 position of guanine, and N1 and N3 positions of pseudouridine [38]. Similarly, kethoxal, which forms adducts with N1 and N2 of guanine, has also been used to probe secondary structure (Table 2.1) [20, 37, 43].

Another important observation is that RNA bases can be protected from chemical modification by interactions with RBPs *in vitro* [34, 40]. Notably, the interactions of several ribosomal proteins with 16S rRNA were heavily scrutinized *in vitro* using DMS and RNase V1 [44–49] with specific interactions corroborated by *in vivo* DMS probing [39].

More recently, chemical based structure probing has been adapted to study folding of RNAs within living cells, which are permeable or can be permeabilized to allow diffusion of these molecules into cells. DMS in particular can easily permeate throughout intact cells [42]. As such, these methods have been used in several organisms to map unpaired, solvent-accessible nucleotides *in vivo* [35–37, 41, 50] Notably, in many cases, inferences of RNA folding from *in vivo* structure probing diverged from *in vitro* probing and *in silico* predictions.

2.3 Backbone Modification-Based Methods

The structural consequences of RNA base pairing extend beyond the bases involved and can affect the local geometry of other chemical groups on the RNA molecule. It is this property that allows the use of chemical probes that target the ribose backbone of RNA to determine its secondary structure (Fig. 2.1).

One key observation has been that the reactive 2′ hydroxyl group of each ribose is vulnerable to attack by electrophiles in some structural contexts, but protected in others. When the attached base is involved in a pairing interaction, the geometry of the sugar is constrained and its 2′ OH is blocked by the adjacent phosphodiester. When the base is unpaired, the conformation is more flexible and passes through a wider range of positions, some of which favor electrophilic attack [51].

This property has been exploited to explore RNA secondary structure by selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE). Like other chemical modification techniques, SHAPE involves modification of the RNA molecule in a way that interferes with the activity of RT resulting in incomplete reverse transcription and DNA fragments of different sizes. The locations of RT stops can be determined by analyzing the lengths of these fragments. An RT stop indicates a base that was unpaired leaving its 2′ OH exposed. Paired bases result in RT read through [51, 52].

It is worth noting that RNA can be constrained, not only by canonical Watson-Crick base pairing, but also by wobble interactions and by more complex structures involving three or more bases (e.g. pseudoknots). SHAPE chemistry is sensitive to these interactions making it possible to use SHAPE to study them, but also adding a level of complexity to the interpretation. This is due to exotic conformations that can actually increase the reactivity of a nucleotide by constraining it to a highly exposed conformation. The most highly reactive positions are often left out of the analysis for this reason (Fig. 2.3).

An important difference between SHAPE and protocols that rely on nucleoside base modification is that SHAPE reagents modify the ribose and not the nucleoside, thus they react identically with all four nucleotides. Artifacts caused by differential reactivity between bases are eliminated along with missing information due to bases that cannot be labeled by a particular reagent.

SHAPE has been actively developed and has seen several improvements and variations since its introduction. It was originally shown to determine secondary structural features of a single, carefully designed, *in vitro* transcribed RNA molecule. Modifications to the priming strategy have generalized the protocol to work with a broader range of transcripts [53], including transcripts purified from cells [54].

Several different reagents have been successfully used as electrophiles for SHAPE chemistry, the first of which was N-methylisotoic anhydride (NMIA). NMIA fulfills all of the requirements for SHAPE, it labels all four nucleotides generically and it causes RT to stall, however the labeling reaction takes tens of minutes to complete [51]. Faster labeling has been achieved using 1-methyl-7-nitroisatoic anhydride (1M7), which reacts with RNA in tens of seconds [55]. The only disadvantage of this reagent is that it is currently not commercially available. The fastest reacting electrophile that has been used is benzoyl cyanide (BzCN). It reacts with RNA fast enough to capture changes in structure over short time courses [54]. However, its high reactivity can make it more challenging to use than 1M7 [56].

Although all SHAPE reagents work in the same general way, they do have differential preferences for particular nucleotide geometries. 1M7 and NMIA have very similar, but not identical, patterns of reactivity with structured RNAs. Additionally, they seem to be most different for bases involved in non-canonical and tertiary interactions. This observation opens up the possibility of studying such interactions by comparing the reactivity profiles of multiple SHAPE reagents on a transcript of interest [57].

3 Nuclease-Based High-Throughput Approaches

In 2010, three groups independently published the first experimental, genome-wide RNA structure analysis using the combination of nuclease digestion and HTS [58–60]. Each report varied in choice of organism, starting RNA pool, nucleases, and

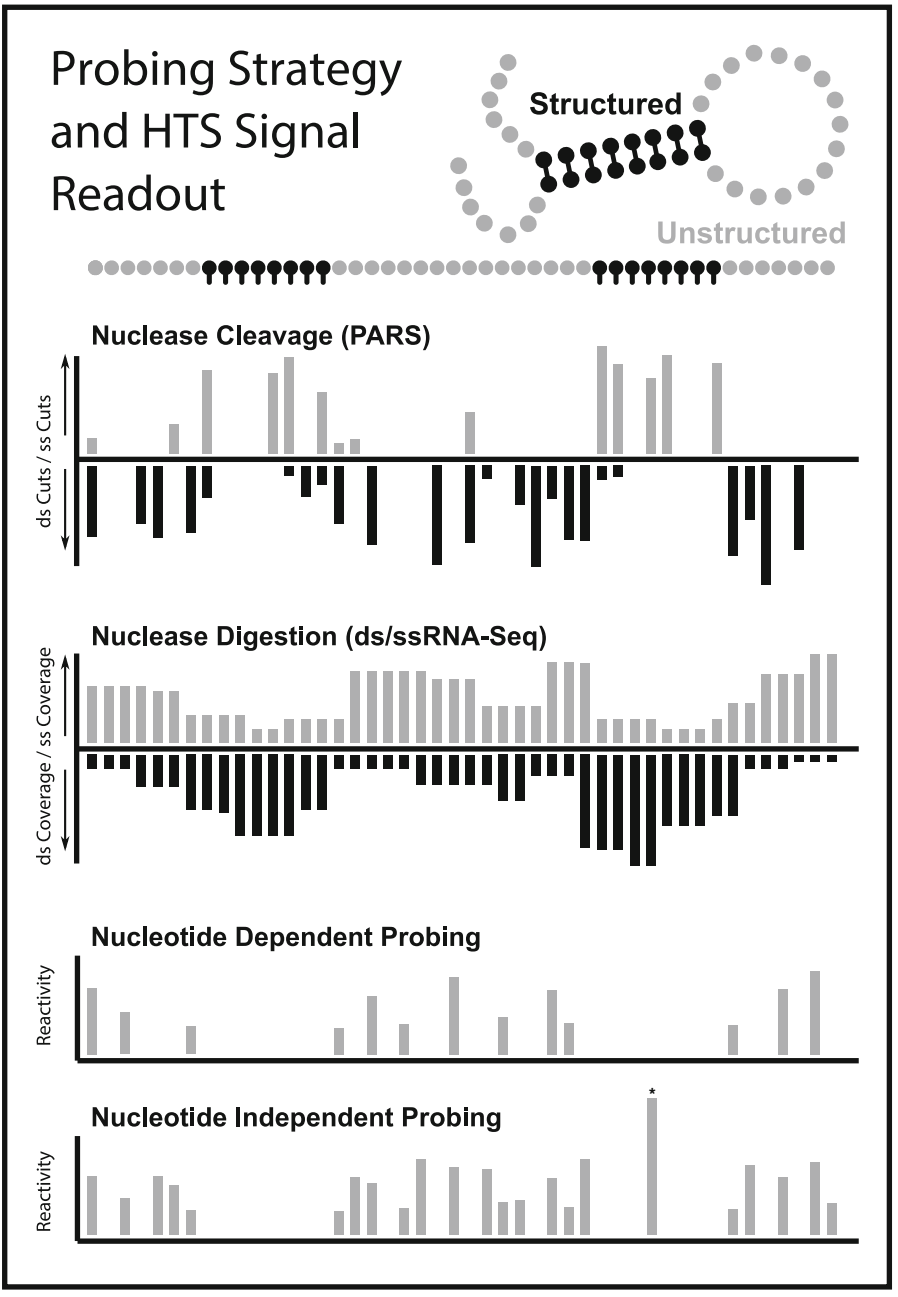


Fig. 2.3 HTS readout signatures of various probe strategies. In PARS, RNases cleave with single hit kinetics resulting in sharp peaks. Signal comes from the region cleaved. The opposite is true for ds/ssRNA-Seq. RNases cleave to near completion and signal comes from the protected regions. The whole read is used resulting in a smooth signal. Chemical probes label unpaired nucleosides resulting in peaks at unstructured regions. The probes are nucleoside specific, so only some positions are substrates. SHAPE reagents label the backbone of RNA at any unstructured nucleotide. However, some positions are structured, but fixed in a highly reactive configuration (denoted with an asterisk in *bottom panel*)

structure calculation approach. Since their introduction, these protocols have been used to study a variety of organisms and RNA conditions. Here, we discuss the basis for these methods and applications thereof (Table 2.2).

3.1 Parallel Analysis of RNA Structures (PARS)

Parallel analysis of RNA structures (PARS) utilizes the structure-specific nature of RNase V1 and RNase S1 for dsRNA and single-stranded RNA (ssRNA) cleavage, respectively [58]. The choice of these enzymes is optimal because, both result in a 5′ P and 3′ OH, making them direct substrates for adapter ligation and HTS (Table 2.1). This approach was developed to interrogate 'single-hit' kinetics of the RNases, which eliminates the possibility of conformational shifts due to free energy changes imparted by enzymatic cleavage. RNase concentrations are optimized such that 10–20 % of RNA molecules are cleaved in 15 min [61]. To allow for HTS, random RNA fragmentation is performed, via metal-mediated hydrolysis to an average size of ~200 nucleotides (nt). Importantly, PNK treatment is not used, such that randomly fragmented RNAs are not substrates for 5′ adapter ligation, thereby selecting for only the 5′ P ends generated by RNase cleavage. Computational analysis is performed by checking base quality, mapping to the transcriptome, and then cleavage sites are determined to be 1 nt upstream of the mapped 5′ end of each sequencing read (Fig. 2.3). Finally, the $\log_2$ ratio of RNase V1/RNase S1 cleavage sites is used to calculate the PARS score.

3.1.1 PARS Analysis in *Saccharomyces cerevisiae*

PARS was first used to study RNA secondary structure on polyA$^+$ RNAs from *Saccharomyces cerevisiae* [58]. Several replicate experiments were performed for each RNase and good correlations were found across replicate experiments (0.6–0.9). In total, ~85 million reads were obtained with a coverage of one read/base for roughly 3000 *S. cerevisiae* mRNA transcripts.

To validate that PARS accurately captures known RNA structures, *S. cerevisiae* polyA$^+$ RNA was spiked with domains from the non-coding RNA HOTAIR, and from the well characterized Tetrahymena group I intron ribozyme. Single RNA footprinting experiments on these same RNAs showed good agreement (correlation = 0.40 – 0.97) with PARS, suggesting that the methodology is able to maintain accuracy in a complex mixture of RNAs. Additionally, PARS data were compared to known RNA structures, including three structural domains from the ASH1 mRNA, an element in URE2 mRNA, and the glutamate transfer RNA. Again, generally good agreement was observed between low- and high-throughput approaches supporting the utility of PARS. Finally, PARS results were compared to computationally predicted structures from RNAfold [15]. A significant correspondence was observed between significantly high PARS

Table 2.2 Genome-wide techniques for RNA structure probing

Method	Probe(s)	Structural specificity	Nucleotide specificity	Informative positions	Organisms studied	Cellular context	Citations
PARS and PARTE	Nuclease S1 and RNase V1	ds and ssRNA	None	1st	Yeast, human	*In vitro* and native deproteinized	[58, 61–63]
FragSeq ds/ ssRNA-seq	Nuclease P1	ssRNA	None	1st and last all	Mouse	*In vitro*	[59]
	RNase I and RNase V1	ds and ssRNA	None		Plants, worms, flies	*In vitro* and native deproteinized	[64–66]
DMS-seq	DMS	ssRNA	A,C only	1st	Yeast, human	*In vitro* and *in vivo*	[71]
Structure-seq	DMS	ssRNA	A,C only	1st	Plants	*In vitro* and *in vivo*	[72]
CIRS-seq	DMS and CMCT	ssRNA	None	1st	Mouse	Native deproteinized	[73]
Mod-seq	DMS	ssRNA	A,C only	1st	Yeast	*In vivo*	[70]
MAP-seq	DMS, CMCT and 1M7	ssRNA	None	1st	Synthetic RNA	N/A	[80]
SHAPE-seq 1.0 and 2.0	NMIA, 1M7, BzCN	ssRNA	None	1st	Synthetic RNA	N/A	[53, 56, 79]

scores and high base-pairing probabilities, and vice versa. However, numerous differences were found between predicted and measured, and thus it was proposed that PARS scores could be used to constrain RNA folding algorithms for a more accurate analysis of RNA secondary structure across eukaryotic transcriptomes.

This study was the first to experimentally examine the structural 'profile' of mRNAs. From this analysis, it was found that *S. cerevisiae* coding regions were on average more highly structured than untranslated regions (UTRs). Furthermore, start and stop codons exhibited "dips" in RNA secondary structure. Intriguingly, a three-nucleotide periodicity in RNA structure across coding sequence (CDS) was identified, with the first nucleotide being least structured and second nucleotide being most structured. The degree of periodicity was associated with translation rates, with high periodicity associated with increased translation and vice versa, suggesting that this structural dynamic is functional. The relationship between overall structure near the start codon and efficiency of translation was also examined. This analysis revealed a very slight (−0.1) negative correlation between structure in this region and translation efficiency, suggesting that in part, structure proximal to the start codon negatively influences translation rates. Gene ontology analysis of transcripts with similar PARS scores identified increased structure in coding regions at genes involved in protein localization or metabolic pathways, while low structure was observed in the 5′ UTR and CDS of mRNAs encoding ribosomal subunits. Finally, transcripts encoding signal peptides were reported to show decreased RNA structure in their 5′ UTRs, which suggests that these unstructured regions allow for interactions with other factors required for their localization.

3.1.2 PARS with Temperature Elevation

This original study was expanded on by applying PARS at different temperatures to reveal the folding energies of RNA structures [62]. In this experiment, termed Parallel Analysis of RNA Structures with Temperature Elevation (PARTE, polyA$^+$ RNAs from yeast were again spiked with specific domains of the exogenous RNAs; HOTAIR, HOTTIP, and *Tetrahymena* ribozyme. For this study, only RNase V1 was used, as the authors were looking for decreases in RNA secondary structure as temperatures were elevated. Three million raw sequencing reads were obtained, which was sufficient to profile structural changes for ~350,000 bases in ~4000 transcripts at various temperatures (23 °C, 30 °C, 37 °C, 55 °C, and 75 °C). Interestingly, approximately 80 % of bases changed from double to single stranded within the temperatures measured.

To validate that the approach was truly measuring changes in RNA secondary structure in response to increased temperatures, melting temperatures (Tms) were measured for 12 specific transcripts using UV spectroscopy. Good concordance was found between this gold standard approach and PARTE measurements ($r = 0.59$). Not surprisingly, higher coverage transcripts were even more strongly correlated

(r = 0.9). PARTE Tms were somewhat correlated with RNAfold (a free energy based structure prediction algorithm) estimates (r = 0.24). However, RNAfold predictions demonstrated a negative correlation with UV spectroscopy, suggesting that PARTE data is more accurate at calculating Tm than RNAfold.

PARTE was easily able to distinguish between coding and noncoding RNAs, where average Tms were higher for canonical structured ncRNAs. Interestingly, UTRs demonstrated intermediate Tms, between the CDS and ncRNAs. However, UTRs have lower PARS scores (less structured) than the CDS, and thus the intermediate Tms of UTRs are most likely a consequence of strong structural elements contained within the mostly unstructured landscape of these transcript regions. The weakest pairing in mRNAs was observed at the start codon, while the most stable region was found to be the area surrounding the stop codon. This was somewhat surprising given that both regions have structural 'dips', and the start codon is also flanked by highly structured regions. Finally, in concordance with what was described in the previous study, transcripts encoding ribosomal protein subunits had low melting temperatures.

3.1.3 PARS Analysis of the Human Transcriptome, *In Vitro* and Native Deproteinized

The most recent study using PARS, applied this approach to RNA isolated from lymphoblastoid cells from a family trio on both denatured polyA$^+$ RNA and rRNA-depleted native deproteinized RNA [63]. This was the first nuclease-based *in vitro* and *in vivo* study in human cells producing structural information for >20,000 transcripts with greater than one read/base. The approaches were validated by comparison to nuclease probing of snoRNA74A and the P9–9.2 domain of *Tetrahymena* ribozyme, two known structures. A method for *in vivo*-like structure analysis was developed by deproteinizing RNA under native conditions, and validated with snoRNA74A. This approach captured structures for ~6000 transcripts. Similar to previous studies in yeast, 'dips' in PARS scores were observed at both the start and stop codons. However, in contrast to yeast, the human CDS was slightly less structured than UTRs, but these differences were not significant. Similar to previous studies, a three nt periodicity was observed in the CDS only, and results were similar between native and renatured RNAs. In fact, by comparing renatured and native regions, they found only ~600 regions with consistently different structures.

An examination of splice sites found that on average, the 3′ end of exons are unstructured and accessible, while the downstream exons are more structured. This analysis also revealed that ARGONAUTE (AGO) binding sites are more unstructured at positions −1–3 nucleotides upstream and at 4–6 of the miRNA binding region compared to untargeted sites. Furthermore, an AGO-CLIP experiment demonstrated that single-stranded miRNA target sites are more bound than structured target sites. They validated this finding by ectopically expressing miRNAs, which were more effective at unstructured targets, indicating that miRNA accessibility is a key determinant of target interaction.

Finally, by comparing structures of three related individuals, single nucleotide variants (SNVs) that modified specific structures in mRNAs, termed riboSNitches were identified and validated using SHAPE or DMS. Excitingly, 15 % of SNVs switched RNA structure in the trio. Numerous riboSNitches overlapped with expression quantitative trait loci (eQTL) and genome-wide association study (GWAS) sites, suggesting a potential link between structural changes and disease. Interestingly, riboSNitches were depleted in 3′ UTRs and around miRNA target sites and RBP binding sites, likely because mutations that disrupt these interactions will be selected against during evolutionary events.

3.2 FragSeq

Fragmentation sequencing (FragSeq) is similar to PARS, except that it relies only on the ssRNase nuclease P1 for identification of unpaired regions [59]. RNase P1 cleaves ssDNA and ssRNA to yield 5′ P and 3′ OH, making them direct substrates for adapter ligation (Table 2.1). In this protocol, RNA is treated with nuclease for 1 h, single-hit kinetics were not considered and therefore restructuring of RNA after cleavage is possible (Table 2.2). However, this is not a significant concern as it is highly unlikely that dsRNA regions would become unstructured after cleavage. This protocol avoids fragmentation by focusing on shorter RNA species. Furthermore, to control for endogenous cleavage events resulting in 5′ P RNA, a set of controls without nuclease and with or without PNK are used. This approach allows the calculation of cleavage probabilities in the nuclease versus control samples, and uses this log ratio to calculate Cutting Scores. Specifically, ssRNA regions are considered to be 1 nt upstream of the 5′ ends and at the 3′ end of trimmed reads.

3.2.1 FragSeq on Mouse Nuclear RNA

In the original report, this method was applied to study *in vitro* refolded, mouse nuclear RNA [59]. Nuclear RNAs are enriched for small to medium sized (70–300 nucelotides) RNAs, which after nuclease treatment results in 20–100 nucleotide RNAs that are ideal size for HTS approaches. To confirm that nuclease P1 activity would not be altered in complex RNA mixtures, nuclear RNA was spiked with the U1a snRNA or 5S rRNA and classical nuclease probing with S1 nuclease was performed.

To confirm that this methodology was working properly and reproducibly, cutting scores for well-described RNAs were compared between replicates. This analysis revealed high Pearson correlation coefficients (0.813–0.889) for several RNAs, indicating reproducibility. They then compared cutting scores for well-described RNA structures U1A, U3b, and U5. Consistent with the known activity of nuclease P1, loops and stem loops of 4–6 bases were easily identifiable, but this approach had low accuracy for small bulges or interior loops. It is noteworthy that since the original report of FragSeq in 2010, there have been no follow up studies, likely due to its similarity to PARS.

3.3 ds/ssRNA-Seq

dsRNA-seq was first described in 2010 [60], and was eventually paired with ssRNA-seq in 2012 [64, 65]. This method relies on the use of a dsRNase (RNase V1) and ssRNase (RNase I) treatment to degrade nuclease-sensitive regions of RNAs in an effort to enrich ssRNA and dsRNA fragments, respectively (Table 2.2). Digested RNA is then subjected to random fragmentation and because RNase I results in 5′ OH and 3′ P, PNK treatment is required for adapter ligation (Table 2.1). Thus, in contrast to PARS and FragSeq, ds/ssRNA-seq does not look for direct cleavage sites but rather aims to sequence the RNA that is remaining after RNA is thoroughly digested with RNase (Fig. 2.3). Briefly, trimmed reads from dsRNA-seq and ssRNA-seq are used to calculate a structure score, defined as a generalized log-odds ratio of dsRNA-/ssRNA-seq reads. This approach lacks some resolution, but allows for a more regional look at RNA structure through sensitivity to RNases and is easily adapted to provide a concomitant look at both RNA secondary structure and RNA-protein interaction sites [66–68].

3.3.1 dsRNA-Seq in *Arabidopsis thaliana* RDR6 Mutants

dsRNA-seq was first used to investigate dsRNAs in rRNA-depleted total RNA from unopened flower buds of *Arabidopsis thaliana* (*Arabidopsis*) [60]. To do this, refolded RNA was subjected to RNase I treatment, enriching for dsRNAs. Indeed, global analysis of tRNAs found that base-paired stem regions had a higher level of dsRNA-seq reads than the unpaired anticodon loop and amino acid acceptor stem. Furthermore, dsRNA-seq reads were strongly biased towards structural RNA molecules, indicating an enrichment for dsRNA species. Interestingly, by investigating sense versus antisense RNAs in dsRNA-seq, it was found that only 16 % of reads sequenced were intramolecular, whereas most were intermolecular. However, many RNA classes were enriched for intramolecular reads, including most classes of functional non-coding RNAs, whereas transposable elements had a strong antisense bias, likely due to endogenous mechanisms that generate dsRNAs through the action of RNA-dependent RNA polymerases (RDRs). Within mRNAs, the 5′ UTR and CDS were biased towards sense (intramolecular) interactions. However, only a slight sense enrichment was observed for 3′ UTRs, likely due to the fact that many 3′ ends of mRNA-encoding genes are highly overlapping in the compact *Arabidopsis* genome.

The specificity of this approach was also leveraged to identify dsRNAs that were specifically generated by RDR6, by performing dsRNA-seq in the presence and absence of this protein. Specifically, using a sliding window approach ~7000 regions were identified where dsRNAs were depleted in *rdr6* mutant compared to wild type plants. A strong internal validation was that seven out of eight previously characterized RDR6 substrates were included in this list. In *Arabidopsis*, many dsRNAs are processed into smRNAs by DICER-LIKE (DCL) proteins, and are incorporated into one of the ten AGO proteins. Therefore, smRNA-seq was also

performed in wild type and *rdr6* mutant plants. Using the same sliding window approach, 218 regions were identified where dsRNA and smRNA were both significantly decreased in the *rdr6* mutant as compared to the wild type control. Interestingly, mRNAs that were found to be RDR6 substrates were enriched for functions in translation and RNA processing.

dsRNA hotspots were also first identified in this study, and defined as regions that were significantly enriched for dsRNA reads as compared to the transcriptome-wide average. Interestingly, transposable elements and mRNAs were the largest source of hotspots. In fact, approximately 2000 mRNAs contained dsRNA hotspots and were enriched for translation and nucleic acid binding. Hotspots were found to be more evolutionarily conserved than flanking sequences in UTRs and introns, but not within the CDS. This approach was also used to identify novel structured RNAs. Finally, a folding algorithm was constrained with the dsRNA-seq data to produce models of secondary structure for all detectable mRNAs.

3.3.2 ds/ssRNA-Seq in *Drosophila melanogaster* and *Caenorhabditis elegans*

ds/ssRNA-seq was then applied to study RNA secondary structure on rRNA-depleted total RNA isolated from *Drosophila melanogaster* (*Drosophila*) DL1 culture cells and *Caenorhabditis elegans* mixed stage worms [64]. This study used an alternative analysis approach by identifying genomic regions that were significantly enriched, either in paired (dsRNA) or unpaired (ssRNA) experiments. 25,000 and 10,000 dsRNA hotpots and 20,000 and 7000 ssRNA hotspots were identified in *Drosophila* and *C. elegans*, respectively. dsRNA hotspots were validated by treating total RNA with structure-specific RNases, followed by RT-PCR amplification. In both organisms, it was revealed that ssRNA hotspots are enriched in protein-coding regions, whereas dsRNA hotspots were predominantly found in transposable elements. Similar to what was described in the previous study, highly structured regions of multiple classes of RNAs were processed into smRNAs. In flies, dsRNA hotspots were enriched at repressive histone marks (H3K9me3), while ssRNA hotspots were enriched at active euchromatin, as indicated by enrichment in H3K4me3 and H3K9ac. Furthermore, smRNA generating hotspots were more highly enriched at the repressive H3K9me3. These results did not hold true in *C. elegans* likely because they don't have the same endogenous siRNA mechanisms. Finally, this study was used to identify previously unannotated regions of both genomes.

ds/ssRNA-seq data were used to constrain RNAfold to generate experimentally derived structural models. Using these structure calls, this feature was interrogated across mRNAs. This analysis revealed strong 'dips' in structure scores at the start and stop codons for both organisms. However, unlike studies in yeast, it was observed that the CDS had significantly lower overall structure than the UTR in both *Drosophila* and *C. elegans*, which is similar to what was observed by PARS analysis of human RNAs and may be a general trend for metazoans [63].

Interestingly, it was also observed that miRNA seed sites were significantly more structured in *Drosophila* but unstructured in *C. elegans*. Furthermore, in *C. elegans*, ALG-1 (the miRNA binding AGO protein) sites were more unstructured at the seed region. Additionally, it was noticed there was a negative correlation between ALG-1 CLIP tag density and structure score, indicating that structure negatively impacts ALG-1 binding to target RNAs.

3.3.3 ds/ssRNA-Seq in *Arabidopsis*

The ds/ssRNA-seq methodology was also applied to study RNA secondary structure in the transcriptome of *Arabidopsis*, in combination with numerous other transcriptome-wide datasets [65]. Using ds/ssRNA-seq data, RNAfold was constrained to generate structural models for thousands of *Arabidopsis* mRNAs. mRNAs were stratified by their overall structure scores and gene ontology (GO) analysis was performed on the 10 % most and least structured transcripts. Intriguingly, highly structured mRNAs were enriched for immune related processes, including cell killing and defense response. In contrast, lowly structured mRNAs were enriched for basic cellular processes, such as transcription, RNA metabolism, and signaling pathways. The low structure of mRNAs involved in basic cellular processes is supported by PARS studies in yeast, which found that mRNAs encoding ribosomal subunits had low structure [58].

By using a peak-finding approach, ~65,000 regions were identified as being significantly structured (hotspots) or unstructured (coldspots). Similar to previous studies in *Arabidopsis* using only dsRNA-seq, transposable elements were enriched for structure hotspots. In contrast, mRNA regions were enriched in coldspots, although there were also a significant number of hotspots in these regions. Also consistent with the previous study in *Arabidopsis*, hotspots and coldspots are more conserved than flanking regions, suggesting their functionality. Furthermore, hotspots were enriched for histone modifications consistent with repressive heterochromatin, including H3K9me2, H3K27me1, and 5mC. In contrast, coldspots were enriched for activating, euchromatic histone modifications, including H3K36me3, H3K4me2, and H3K4me3. This supports the link between RNA secondary structure and genomic repression in plants, which functions through dsRNA-dependent smRNA pathways.

The secondary structure score profile around mRNAs was also examined. Just as the previous studies in yeast, worms, and flies, this study revealed strong dips in structure at the start and stop codons. Similar to what was observed for yeast, but in contrast to worms and flies, the CDS was more highly structured than UTRs. miRNA seed sites were also more accessible, which was previously observed in worms but not flies [58, 64].

When average structure scores were compared to the abundance of RNAs a significant anticorrelation ($r = -0.46$) was observed. To further address this observation, the relationship between RNA structure and rates of degradation, as measured by degradome sequencing, also known as genome-wide mapping of uncapped and

cleaved transcripts (GMUCT) [69], was tested. From this analysis, a significant but slight positive correlation (r = 0.21) was observed, suggesting that more structured mRNAs are turned over more rapidly to regulate their abundance. Furthermore, when examining the relationship between smRNA abundance and mRNA secondary structure a very strong positive correlation (r = 0.62) was observed, supporting the notion that highly structured RNAs may be directly processed into smRNAs to regulate their abundance by leading to their increased rate of degradation. Finally, the correlation between structure score and ribosome occupancy was interrogated. This analysis revealed a significant positive correlation (r = 0.37) between these parameters, suggesting that more structured RNAs are more associated with the ribosome, which could be due to ribosome stalling or increased translation initiation rates. However, the former model was preferred given the high prevalence of smRNA processing and turnover observed for highly structured mRNAs in *Arabidopsis*.

3.3.4 Native Deproteinized ds/ssRNA-Seq in *Arabidopsis*

The most recent ds/ssRNA-seq study used a modified approach, allowing for *in vivo* analysis on total nuclear RNA from *Arabidopsis* [66]. To obtain natively folded RNAs, nuclei were first crosslinked with formaldehyde and treated with low concentrations of SDS and Proteinase K to disrupt and remove bound proteins. In addition to analyzing RNA secondary structure, the authors also profiled RBP interaction sites using a global RBP footprinting approach, originally developed in HeLa cells and known as protein interaction profile sequencing (PIP-seq) [68]. Interestingly, it was found that RBP occupancy and secondary structure were anticorrelated at every region investigated. This included the CDS start and stop sites as well as 3′ and 5′ splice sites. Additionally, various types of alternative splicing events (i.e., exon skipping, intron retention, and U12-dependent) were found to have distinct structural and RBP occupancy profiles. Furthermore, it was found that bound instances of RBP binding motifs were less structured than their unbound counterparts, again supporting the notion that RNA structure and RBP binding sites are anticorrelated.

4 Chemical-Based High-Throughput Approaches

4.1 Base Modification-Based High-Throughput Approaches

DMS, which can permeate cellular membranes, has been particularly useful for the elucidation of RNA structure *in vivo* in various organisms [36]. These primer extension-based methods allowed probing across long molecules, and facilitated robust inferences of functional RNA structures in tRNAs, rRNAs, and introns.

However, this approach was limited by the need for custom oligonucleotides to probe any given region and restricted to abundant transcripts. These limitations were addressed in 2014, when three studies (two in yeast, one in human, and one in *Arabidopsis*) integrated DMS structure probing into next-generation RNA sequencing (RNA-seq) pipelines, enabling parallel assessment of secondary structure in thousands of transcripts (Table 2.2).

For *in vivo* analysis, intact cells or plants were treated with DMS prior to RNA purification. In one study of yeast and human RNA secondary structure, DMS reactivity was also measured in denatured and *in vitro* refolded RNA. If desired, specific RNA classes can be enriched using polyA$^+$ selection or rRNA-depletion, facilitating the analysis of lower abundance transcripts. Isolated RNA is then fragmented and HTS libraries are generated using a variety of second strand cDNA sequencing strategies. In two studies, DMS-independent RT dropoff was controlled for by generating HTS libraries similarly to the *in vivo* samples excluding DMS treatment.

While these techniques provide nucleotide resolution reactivity data, there are a number of caveats associated with these approach. For instance, DMS only reacts with As and Cs, G and U positions are uninformative (Fig. 2.1). Moreover, as with all chemical structure probing techniques, DMS-seq does not provide complementary information regarding propensity of a nucleotide to be in the paired state, therefore, regions of RNA with low DMS reactivity are not necessarily structured [40, 70]. Furthermore, positions with a high natural propensity for RT drop-off, such as endogenous nucleobase modifications and exceptionally stable secondary structures can confound individual reactivity measurements in the absence of appropriate controls [20].

4.1.1 DMS-Seq in Yeast and Human Cells

Recently, DMS-seq was applied to yeast, K562 cells, and human foreskin fibroblasts [71]. RNA was treated with DMS *in vivo*, *in vitro*, and after denaturation and in all cases polyA$^+$ selection was used to enrich for mRNA (Table 2.2). Consistent results were observed between experiments in yeast even with substantial variations in DMS treatment conditions, suggesting the robustness of the assay. These yeast libraries were sequenced to significant depth, with >140 million uniquely mapping reads in each sample. Analysis of 3 validated yeast mRNA structural elements as well the 18S and 25S rRNA found that DMS reactivity can be used to resolve structure with nucleotide resolution and up to 94 % accuracy (18S rRNA).

DMS reactivity was also used to assess trends in mRNA structure after *in vivo* and *in vitro* probing. This analysis was restricted to mRNAs for which A and C positions were supported by at least 15 reads on average. Rather than directly calculate reactivity scores per nucleotide for structural analysis, mRNAs were segmented into bins with a constant number of A and C positions. For each bin, correlation between and evenness in DMS reactivity of the structured sample and denatured control were calculated. Therefore, bins that showed low correlation and

evenness of DMS reactivity in the structured sample compared to the control were regarded as structured.

A sampling of 5000 random bins showed that *in vivo* 3.9 % and 29 % of tested regions show structure or are indistinguishable from the denatured control, respectively. Conversely, by this same assessment, *in vitro* probing indicated 24 % and 9 % of regions are structured and highly unstructured, respectively. Notably, by this analysis, *in vivo* probing defined all 50 tested A/C bins corresponding to rRNA and validated mRNA structures as structured. However, although *in vitro* probing accurately classified validated mRNA structures, only ~12 % of tested rRNA regions were called correctly as being structured in nature. This study also reports that average structure is similar between the UTRs and CDS, *in vivo*, which is in contrast to PARS data from yeast [58].

While this analysis approach can help address the 3′ bias of mRNA-seq, correlation and relative evenness of DMS reactivity within bins with respect to a denatured control may be sensitive to changes in bin size. For instance, although contraction of bins results in minimal changes in fractions of regions classified as structured and unstructured, increasing bins to 100 A/Cs approximately doubles the number of tested bins which are defined as structured. Moreover, this improves the accuracy of rRNA categorization during *in vitro* probing to around 50 %. However, the trend of higher structure inferred by *in vitro* probing holds true regardless of bin size. This, with the observation that *in vivo* probed structure increases in ATP-depleted yeast, supports the hypothesis that structured regions of mRNA are actively unwound within normal cells, which the authors speculate is due to active unwinding by RNA helicases. However, it is also probable that this observation is due to active unwinding of RNAs by ribosomes during translation.

4.1.2 Structure-Seq in *Arabidopsis*

Concurrently, an analogous approach to DMS-seq, termed structure-seq, was performed on *Arabidopsis* seedlings [72]. The primary difference between these two approaches is that for the *Arabidopsis* study detection of RT stops in the absence of DMS treatment was used as the background control. It is worth noting that this method appropriately controls for potential RT stops due to naturally occurring nucleotide modifications and particularly stable RNA structures, while the DMS-seq uses denatured RNA with DMS as a control (Table 2.2) [20, 71].

In this study, DMS reactivity in *Arabidopsis* was compared to a phylogenetically derived model of 18S rRNA secondary structure. Interestingly, nucleotides of high reactivity were largely predicted to be unpaired through conservation analysis, while nucleotides of low reactivity were almost evenly distributed between phylogenetically predicted paired and unpaired regions. While these “false negatives” for DMS reactivity might be due to binding of ribosomal proteins or RNA tertiary structure, it is possible that these samples were not sequenced at enough depth (200 million reads total for this analysis, as opposed to 200 million per

sample) to achieve the same level of accuracy observed in the DMS-seq experiment for yeast.

Meta-analysis of DMS reactivity in mRNAs indicated that structure is modestly, but significantly higher in the CDS compared to the UTRs, which mirrors previous *in vitro* observations of total cellular *Arabidopsis* mRNA, but contradicts data from natively folded mRNA in the nucleus [65, 66]. Additionally, highly translated transcripts were demarcated by a spike in DMS reactivity upstream of the start codon and a characteristic 3 nt periodicity in reactivity, which were absent in both lowly translated mRNAs and in the UTRs. The activity near the start codon corroborates other evidence of a dip in secondary structure at the start codon [58, 62–66, 73]. Moreover, on average the first codon nt was found to be the least structured while the second codon position was the most structured, which matches results from the *in vitro* PARS analysis of yeast [58], and may indicate a highly conserved mechanism of translational regulation.

This study also compared *in silico* predictions of mRNA folding that were either unconstrained or constrained by *in vivo* or *in vitro* structure data. This analysis suggested that *in vivo* constraints caused predictions to be similarly disparate from unconstrained and *in vitro* constrained mRNA folding predictions. In total, these findings led to the speculation that this observation is not due to RBP binding because of a negative correlation between DMS reactivity and the similarity between *in vivo* constrained and unconstrained positions. However, DMS-seq is sensitive to RBP binding, and thus further experimentation is required to deconvolute these observations [40, 70]. Intriguingly, highly discordant predictions tended to include transcripts associated with stress response, while those associated with critical cellular functions were enriched in the most concordant structure predictions.

4.1.3 Mod-Seq in Yeast

An additional DMS-based structure mapping protocol was ,also recently applied to the yeast transcriptome. This iteration, known as Mod-seq (Table 2.2), follows a very similar approach to the *Arabidopsis* study, where RT stops were assessed with and without *in vivo* DMS modification [70]. However, mRNAs were not enriched, and as such, this study focused on DMS reactivity in rRNAs. Additionally, this study was the only one to use replicate information to determine statistical significance for DMS-reactivity assessments compared to controls. Notably, a companion software package known as Mod-seeker was released to automate and standardize this analysis.

Mod-seq of yeast found that positions with >1.5-fold and significant enrichment in DMS induced RT stops are consistent with known unstructured regions of ncRNAs. Mod-seq analysis of a ribosomal protein L26 deletion strain was also compared to wild type to identify the known binding sites of this RBP. Moreover, strong RT stops at known endogenous RNA modifications were detected in the no-treatment controls for this analysis. Together, these findings indicate that while DMS can be used to reliably detect unstructured regions of RNA, both endogenous

RNA modifications and RBP binding can strongly confound the analysis of this approach. Thus, until these problems can be resolved structural models determined by this method are likely to contain a number of inaccurately determined positions throughout their length.

4.1.4 CIRS-Seq in Mouse Embryonic Cells

An additional high-throughput small molecule, probing method for RNA structure determination is known as chemical inference of RNA structures (CIRS-seq) [73]. This method deploys two small molecules, DMS and CMCT, and was recently used to probe natively folded, deproteinated RNA (Table 2.2). This strategy improves resolution as compared to DMS-based approaches alone by allowing any canonical nucleobase in an unstructured confirmation to be probed and depletes potential interference from RBPs through deproteination. As with structure- and Mod-seq, HTS is used to identify sites of enriched RT drop-off due to chemical treatment compared to a non-treated control.

CIRS-seq was applied to mouse embryonic stem cells and rRNA was depleted after chemical treatments. A large percentage of highly reactive positions (>73 %) identified by CIRS-seq within validated tRNA and snRNA structural models are known to be unstructured, which improves further (>87 %) if helix termini that are known to be structurally flexible are excluded from this analysis [74]. CIRS-seq reactivity information was also used to constrain an RNA structure prediction tool and these predictions recapitulate experimentally or phylogenetically defined structures with extremely high accuracy, far surpassing strictly MFE based predictions.

Global analysis of mRNAs also identified many distinct structural features. For instance, distinct, local spikes in chemical modification were observed both immediately preceding the start and at the stop codons, providing further evidence for a highly conserved decrease in secondary structure at these locations. This analysis also indicated that nucleotides in the CDS are, on average, more accessible to chemical modification than those in the UTRs. This is the reverse of the *in vivo* and *in vitro* experiments for total polyA$^+$ RNA from *Arabidopsis*, but similar to studies in other organisms [64–66, 72].

As with PARS and structure-seq, CIRS-seq was able to identify a 3 nt periodicity in reactivity within the CDS [58, 63, 72]. Average reactivity measurements indicate that the first codon position is generally the least structured as was seen in plants and yeast. However, in this study, the third codon position was inferred to be the most structured while previous studies indicated the second position is the most likely to be paired. While this may be due to functional variation between organisms, these disparate observations may also be due to experimental noise.

Finally, CIRS-seq reactivity measurements were used to infer average secondary structure near Lin28a binding sites identified in a previous CLIP-seq experiment. This analysis identified a distinct increase in chemical accessibility at expected Lin28a binding sites. They also predict that these regions tend to form hairpin structures, with Lin28a interacting with the loop region.

4.2 SHAPE-Based High-Throughput Approaches

The evolution of SHAPE into a high-throughput technique has closely followed the evolution of RNA sequencing and has made use of many of the same technologies. Originally, the RNA fragments generated in SHAPE reactions were radiolabeled and resolved using gel electrophoresis, similar to protocols for Sanger sequencing [52].

Just as Sanger sequencing has been adapted into a high-throughput methodology, a high throughput adaptation of SHAPE (i.e., hSHAPE and SHAPE-CE) has been developed. In this protocol, primers are labeled with color-coded fluorophores and resolved using capillary electrophoresis. Contrasting fluorophores allow multiple samples to be multiplexed into a single capillary and separated computationally [75].

This method has been used to probe the structures of RNAs from several viral pathogens. One study determined the structure of a region of the HIV genome *in virio* [75]. Another looked at structural differences and similarities between members of a family of viroids, pathogenic RNAs that parasitize plants [76]. It has also been used to draw a relationship between the structure of a viral internal ribosome entry site and that virus's pathogenicity [77], as well as to study a structured viral sequence involved in translation reinitiation [78].

The next natural step in improving SHAPE throughput has been the combination of SHAPE chemistry with HTS technology (SHAPE-seq) (Table 2.2). This process started when SHAPE-seq was done for *B. subtilis* RNase P, a well characterized catalytic RNA. Its reactivity pattern was shown to match the known structure as well as previous results from SHAPE. This approach was also used to demonstrate the ability to take advantage of the multiplexing capabilities of HTS by performing SHAPE-Seq on 256 individual barcoded clones of RNase P [79]. Another group has proposed a protocol for combining SHAPEwith DMS and CMCT to create multiple marks capable of stopping RT, which could potentially compensate for the biases of each reagent [80].

Recent developments in SHAPE-seq technology have allowed the first unbiased transcriptome-wide interrogation of RNA secondary structure with a SHAPE reagent [81]. Using a novel biotinylatable SHAPE reagent and click chemistry, Spitale et al. successfully affinity purified acylated RNAs after modification, thus reducing background and increasing informative positions. The high accuracy of in vivo click SHAPE (icSHAPE) will likely preclude a transformation in the high-throughput RNA structure probing field, taking it from nascent methods to mature technology. Expanded use of icSHAPE and similar approaches will continue to expand our understanding of the global landscape of RNA structure and likely reveal unexpected findings embedded within the secondary structure of RNAs.

5 Conclusions and Outlook

In this chapter, we have a discussed a diverse array of strategies for genome-wide RNA secondary structure analysis, each with its strengths and caveats (Table 2.2). For a detailed review on the results from these studies, see [82]. Although some conclusions differ, it is promising that such techniques often arrive at similar discoveries. For instance, 'dips' in RNA secondary structure have been observed at the start and/or stop codons for numerous organisms by several methodologies [58, 62–66, 73]. This appears to be a conserved feature from yeast to humans and likely plays a role in translation initiation and termination.

Another trend that is consistent between methodologies and organisms is the likelihood for base-pairing interactions to occur with a fixed periodicity within the CDS. Both enzymatic (PARS) and chemical (DMS-seq, structure-seq, and CIRS-seq) methods detected this pattern. All methods detected the first nucleotide of each codon as the least structured, however, PARS analysis in yeast and humans and structure-seq in *Arabidopsis* show the second nucleotide as the most structured, whereas CIRS-seq suggests that the third base is the most paired [58, 63, 72, 73].

In *Arabidopsis*, a weak positive correlation was observed between overall transcript structure and ribosome-association [65]. Relatedly, high DMS reactivity (low structure) around the start codon or distinct 3 nt periodicity in the CDS were observed for highly translated but not lowly translated mRNAs [72]. In yeast, a similar observation was made [58]; thus RNA structure is a key determinant of translation rates as has been demonstrated by reporter assays [83]. Additionally, mRNAs encoding "housekeeping" genes, such as RNA metabolic, transcription, and rRNA genes tend to have less structure [58, 62, 65]. Functional miRNA binding sites were found to be unstructured in *C. elegans* and humans likely because structures can protect miRNA target sites from binding [63, 64]. Finally, structure and RNA stability were found to be anticorrelated in *Arabidopsis*, and relatedly in yeast, strong structures were found in the 3′ UTRs of genes with longer half lives [62, 65].

One of the most compelling findings from genome-wide studies is the affect of SNPs on RNA secondary structure. These so called riboSNitches, some of which are enriched near eQTL and GWAS loci, raise the intriguing possibility that sequence differences with functional consequences may act through RNA secondary structure rather than directly through RNA or DNA sequence [63].

As these methods continue to develop and evolve, it is important to consider caveats that arise and must be addressed. One key question is how faithfully *in vitro* structure probing recapitulates true *in vivo* RNA biology. For instance, structural determination of biologically produced RNAs by chemical probing *in vivo* introduces several complexities. One is that in a cellular context, RNAs are constantly associated with proteins and RNA-protein interactions are likely to block these chemical probes [40, 70]. Additionally, for all structure probing experiments, one must consider the buffer environment in the cell, as well as the presence of precise

amounts of ions and other small molecules that can affect RNA structure. Furthermore, active mechanisms such as translation and helicase activity can relieve structural elements and strip them of their bound factors [71]. Finally, modified nucleotides such as (1-methyl guanosine (m^1G)) can affect the processivity of the RT reaction, a step shared by all of the techniques we have discussed [84]. However, these technical hurdles can all be addressed with careful design of controls and may even be eliminated with future protocol improvements. It is also noteworthy that care must be taken in design of the analytical approaches and final interpretation of all results from these high-throughput structural studies.

Now that there are so many reagents and approaches, empirical comparisons can be made and the best approaches should become the most widespread. Optimization and standardization of computational pipelines for analysis of genome-wide structure analysis will help to enable more continuity between labs and experiments [85]. Finally, by coupling these analyses with other genome-wide datasets, our understanding of the impacts of structure are likely to expand [58, 63, 65, 72, 73].

As these techniques become more widespread, we expect a rapid increase in the number of organisms studied and cellular states profiled. With the decrease in RNA sequencing costs, these studies can also achieve higher sequencing depth, which is likely to reveal even higher accuracy in secondary structure determination. All of these advances are important because there are countless open questions that remain to be addressed concerning RNA secondary structure and its numerous functions. For instance, what does the evolutionary conservation of RNA secondary structure across species look like? Are their riboswitches in eukaryotes that remain to be discovered? What is the structure of specific lncRNAs, and are there overarching principles that can be used to connect their structure and function? What are the structural rules that govern miRNA targeting? And what is the connection between RNA structure and disease? Answers to these questions, and more that will arise from future studies, will advance our understanding of this remarkable biomolecule and the functional significance of its structure.

Acknowledgments We thank past and present members of the Gregory lab for helpful discussions, especially Qi Zheng, Fan Li, and Lee Vandivier. This work was supported by an NSF Career Award MCB-1053846 and NSF grant MCB-1243947 to BDG. We declare no competing financial interests.

References

1. Cruz JA, Westhof E (2009) The dynamic landscapes of RNA architecture. Cell 136:604–609. doi:10.1016/j.cell.2009.02.003
2. Kim SH (1978) Three-dimensional structure of transfer RNA and its functional implications. Adv Enzymol Relat Areas Mol Biol 46:279–315
3. Yusupova G, Yusupov M (2014) High-resolution structure of the eukaryotic 80S ribosome. Annu Rev Biochem 83:467–486. doi:10.1146/annurev-biochem-060713-035445

4. Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29:11–17. doi:10.1016/j.tibs.2003.11.004
5. Bocobza SE, Aharoni A (2014) Small molecules that interact with RNA: riboswitch-based gene control and its involvement in metabolic regulation in plants and algae. Plant J Cell Mol Biol 79:693–703. doi:10.1111/tpj.12540
6. Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, Klausner RD (1987) Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. Science 238:1570–1573
7. Williams AS, Marzluff WF (1995) The sequence of the stem and flanking sequences at the 3′ end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. Nucleic Acids Res 23:654–662
8. Pelletier J, Sonenberg N (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature 334:320–325. doi:10.1038/334320a0
9. Mercer TR, Mattick JS (2013) Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20:300–307. doi:10.1038/nsmb.2480
10. Novikova IV, Hennelly SP, Sanbonmatsu KY (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res 40:5034–5051. doi:10.1093/nar/gks071
11. Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH, Seeman NC, Rich A (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science 185:435–440
12. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. Nature 250:546–551
13. Holbrook SR, Kim S-H (1997) RNA crystallography. Biopolymers 44:3–21. doi:10.1002/(SICI)1097-0282(1997)44:1<3::AID-BIP2>3.0.CO;2-Z
14. Scott LG, Hennig M (2008) RNA structure determination by NMR. In: Keith JM (ed) Methods in molecular biology™. Humana Press, New York
15. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA websuite. Nucleic Acids Res 36:W70–W74. doi:10.1093/nar/gkn188
16. Mathews DH (2014) RNA secondary structure analysis using RNAstructure. Curr Protoc Bioinforma 46:12.6.1–12.6.25. doi:10.1002/0471250953.bi1206s46
17. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. Nucleic Acids Res 31:439–441
18. Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res 9:133–148
19. Chang SH, RajBhandary UL (1968) Studies on polynucleotides. LXXXI. Yeast phenylalanine transfer ribonucleic acid: partial digestion with pancreatic ribonuclease. J Biol Chem 243:592–597
20. Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B (1987) Probing the structure of RNAs in solution. Nucleic Acids Res 15:9109–9128
21. Loverix S, Steyaert J (2001) Deciphering the mechanism of RNase T1. Methods Enzymol 341:305–323
22. Uchida T, Arima T, Egami F (1970) Specificity of RNase U2. J Biochem (Tokyo) 67:91–102
23. Volkin E, Cohn WE (1953) On the structure of ribonucleic acids. II The products of ribonuclease action. J Biol Chem 205:767–782
24. Desai NA, Shankar V (2003) Single-strand-specific nucleases. FEMS Microbiol Rev 26:457–491
25. Knapp G (1989) Enzymatic approaches to probing of RNA secondary and tertiary structure. Methods Enzymol 180:192–212
26. Silberklang M, Gillum AM, RajBhandary UL (1977) The use of nuclease P1 in sequence analysis of end group labeled RNA. Nucleic Acids Res 4:4091–4108
27. Favorova OO, Fasiolo F, Keith G, Vassilenko SK, Ebel JP (1981) Partial digestion of tRNA—aminoacyl-tRNA synthetase complexes with cobra venom ribonuclease. Biochemistry (Mosc) 20:1006–1011

28. Lockard RE, Kumar A (1981) Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom. Nucleic Acids Res 9:5125–5140
29. Lowman HB, Draper DE (1986) On the recognition of helical RNA by cobra venom V1 nuclease. J Biol Chem 261:5396–5403
30. Nicholson AW (2014) Ribonuclease III mechanisms of double-stranded RNA cleavage. Wiley Interdiscip Rev RNA 5:31–48. doi:10.1002/wrna.1195
31. Peattie DA (1979) Direct chemical method for sequencing RNA. Proc Natl Acad Sci U S A 76:1760–1764
32. Peattie DA, Gilbert W (1980) Chemical probes for higher-order structure in RNA. Proc Natl Acad Sci U S A 77:4679–4682
33. Inoue T, Cech TR (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. Proc Natl Acad Sci U S A 82:648–652
34. Lempereur L, Nicoloso M, Riehl N, Ehresmann C, Ehresmann B, Bachellerie JP (1985) Conformation of yeast 18S rRNA. Direct chemical probing of the 5′ domain in ribosomal subunits and in deproteinized RNA by reverse transcriptase mapping of dimethyl sulfate-accessible. Nucleic Acids Res 13:8339–8357
35. Antal M, Boros É, Solymosy F, Kiss T (2002) Analysis of the structure of human telomerase RNA in vivo. Nucleic Acids Res 30:912–920. doi:10.1093/nar/30.4.912
36. Ares M, Igel AH (1990) Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. Genes Dev 4:2132–2145
37. Harris KA, Crothers DM, Ullu E (1995) In vivo structural analysis of spliced leader RNAs in Trypanosoma brucei and Leptomonas collosoma: a flexible structure that is independent of cap4 methylations. RNA 1:351–362
38. Metz DH, Brown GL (1969) Investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. Reaction between N-cyclohexyl-N′-β-(4-methylmorpholinium)ethylcarbodiimide and model nucleotides. Biochemistry (Mosc) 8:2312–2328. doi:10.1021/bi00834a012
39. Moazed D, Robertson JM, Noller HF (1988) Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. Nature 334:362–364. doi:10.1038/334362a0
40. Tijerina P, Mohr S, Russell R (2007) DMS footprinting of structured RNAs and RNA-protein complexes. Nat Protoc 2:2608–2623. doi:10.1038/nprot.2007.380
41. Wells SE, Hughes JM, Igel AH, Ares M (2000) Use of dimethyl sulfate to probe RNA structure in vivo. Methods Enzymol 318:479–493
42. Lawley PD, Brookes P (1963) Further studies on the alkylation of nucleic acids and their constituent nucleotides. Biochem J 89:127–138
43. Litt M (1969) Structural studies on transfer ribonucleic acid. I. Labeling of exposed guanine sites in yeast phenylalanine transfer ribonucleic acid with kethoxal. Biochemistry 8:3249–3253. doi:10.1021/bi00836a017
44. Powers T, Changchien LM, Craven GR, Noller HF (1988) Probing the assembly of the 3′ major domain of 16S ribosomal RNA. Quaternary interactions involving ribosomal proteins S7, S9 and S19. J Mol Biol 200:309–319
45. Powers T, Stern S, Changchien LM, Noller HF (1988) Probing the assembly of the 3′ major domain of 16S rRNA. Interactions involving ribosomal proteins S2, S3, S10, S13 and S14. J Mol Biol 201:697–716
46. Stern S, Wilson RC, Noller HF (1986) Localization of the binding site for protein S4 on 16S ribosomal RNA by chemical and enzymatic probing and primer extension. J Mol Biol 192:101–110
47. Stern S, Changchien LM, Craven GR, Noller HF (1988) Interaction of proteins S16, S17 and S20 with 16S ribosomal RNA. J Mol Biol 200:291–299
48. Stern S, Powers T, Changchien LM, Noller HF (1988) Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16S rRNA. J Mol Biol 201:683–695

49. Svensson P, Changchien LM, Craven GR, Noller HF (1988) Interaction of ribosomal proteins, S6, S8, S15 and S18 with the central domain of 16S ribosomal RNA. J Mol Biol 200:301–308
50. Zaug AJ, Cech TR (1995) Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. RNA 1:363–374
51. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127:4223–4231. doi:10.1021/ja043822v
52. Wilkinson KA, Merino EJ, Weeks KM (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 1:1610–1616. doi:10.1038/nprot.2006.249
53. Loughrey D, Watters KE, Settle AH, Lucks JB (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res 42:000–000. doi:10.1093/nar/gku909
54. Hector RD, Burlacu E, Aitken S, Bihan TL, Tuijtel M, Zaplatina A, Cook AG, Granneman S (2014) Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. Nucleic Acids Res 42:12138–12154. doi:10.1093/nar/gku815
55. Mortimer SA, Weeks KM (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc 129:4144–4145. doi:10.1021/ja0704028
56. Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB (2012) SHAPE-seq: high-throughput RNA structure analysis. Curr Protoc Chem Biol 4:275–297. doi:10.1002/9780470559277.ch120019
57. Steen K-A, Rice GM, Weeks KM (2012) Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. J Am Chem Soc 134:13160–13163. doi:10.1021/ja304027m
58. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E (2010) Genome-wide measurement of RNA secondary structure in yeast. Nature 467:103–107. doi:10.1038/nature09322
59. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. Nat Methods 7:995–1001. doi:10.1038/nmeth.1529
60. Zheng Q, Ryvkin P, Li F, Dragomir I, Valladares O, Yang J, Cao K, Wang L-S, Gregory BD (2010) Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. PLoS Genet 6, e1001141. doi:10.1371/journal.pgen.1001141
61. Wan Y, Qu K, Ouyang Z, Chang HY (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. Nat Protoc 8:849–869. doi:10.1038/nprot.2013.045
62. Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, Makino DL, Nutter RC, Segal E, Chang HY (2012) Genome-wide measurement of RNA folding energies. Mol Cell 48:169–181. doi:10.1016/j.molcel.2012.08.008
63. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, Chang HY (2014) Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505:706–709. doi:10.1038/nature12946
64. Li F, Zheng Q, Ryvkin P, Dragomir I, Desai Y, Aiyer S, Valladares O, Yang J, Bambina S, Sabin LR, Murray JI, Lamitina T, Raj A, Cherry S, Wang L-S, Gregory BD (2012) Global analysis of RNA secondary structure in two metazoans. Cell Rep 1:69–82. doi:10.1016/j.celrep.2011.10.002
65. Li F, Zheng Q, Vandivier LE, Willmann MR, Chen Y, Gregory BD (2012) Regulatory impact of RNA secondary structure across the Arabidopsis transcriptome. Plant Cell 24:4346–4359. doi:10.1105/tpc.112.104232

66. Gosai SJ, Foley SW, Wang D, Silverman IM, Selamoglu N, Nelson ADL, Beilstein MA, Daldal F, Deal RB, Gregory BD (2015) Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. Mol Cell 57:376–388. doi:10.1016/j.molcel.2014.12.004
67. Silverman IM, Gregory BD (2015) Transcriptome-wide ribonuclease-mediated protein footprinting to identify RNA-protein interaction sites. Methods 72:76–85. doi:10.1016/j.ymeth.2014.10.021
68. Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, Gregory BD (2014) RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol 15:R3. doi:10.1186/gb-2014-15-1-r3
69. Willmann MR, Berkowitz ND, Gregory BD (2014) Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes—GMUCT 2.0. Methods 67:64–73. doi:10.1016/j.ymeth.2013.07.003
70. Talkish J, May G, Lin Y, Woolford JL, McManus CJ (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. RNA 20:713–720. doi:10.1261/rna.042218.113
71. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505:701–705. doi:10.1038/nature12894
72. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505:696–700. doi:10.1038/nature12756
73. Incarnato D, Neri F, Anselmi F, Oliviero S (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. Genome Biol 15:491. doi:10.1186/PREACCEPT-1911964213137914
74. Yu E, Fabris D (2003) Direct probing of RNA structures and RNA-protein interactions in the HIV-1 packaging signal by chemical modification and electrospray ionization fourier transform mass spectrometry. J Mol Biol 330:211–223. doi:10.1016/S0022-2836(03)00589-8
75. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol 6(4), e96. doi:10.1371/journal.pbio.0060096
76. Giguère T, Adkar-Purushothama CR, Bolduc F, Perreault J-P (2014) Elucidation of the structures of all members of the Avsunviroidae family. Mol Plant Pathol 15:767–779
77. García-Nuñez S, Gismondi MI, König G, Berinstein A, Taboga O, Rieder E, Martínez-Salas E, Carrillo E (2014) Enhanced IRES activity by the 3′UTR element determines the virulence of FMDV isolates. Virology 448:303–313. doi:10.1016/j.virol.2013.10.027
78. Gao F, Gulay SP, Kasprzak W, Dinman JD, Shapiro BA, Simon AE (2013) The kissing-loop T-shaped structure translational enhancer of pea enation mosaic virus can bind simultaneously to ribosomes and a 5′ proximal hairpin. J Virol 87:11987–12002. doi:10.1128/JVI.02005-13
79. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP (2011) Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A 108:11063–11068. doi:10.1073/pnas.1106501108
80. Seetin MG, Kladwang W, Bida JP, Das R (2014) Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. Methods Mol Biol 1086:95–117. doi:10.1007/978-1-62703-667-2_6
81. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung JW, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Change HY (2015) Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519(7544): 486–490. doi:10.1038/nature14263
82. Mortimer SA, Kidwell MA, Doudna JA (2014) Insights into RNA structure and function from genome-wide studies. Nat Rev Genet 15:469–479. doi:10.1038/nrg3681

83. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. Science 324:255–258. doi:10.1126/science.1170160
84. Ryvkin P, Leung YY, Silverman IM, Childress M, Valladares O, Dragomir I, Gregory BD, Wang L-S (2013) HAMR: high-throughput annotation of modified ribonucleotides. RNA 19:1684–1692. doi:10.1261/rna.036806.112
85. Aviran S, Pachter L (2014) Rational experiment design for sequencing-based RNA structure mapping. RNA 20:1864–1877. doi:10.1261/rna.043844.113

Chapter 3
Tethered Function Assays as Tools to Elucidate the Molecular Roles of RNA-Binding Proteins

Tomas J. Bos, Julia K. Nussbacher, Stefan Aigner, and Gene W. Yeo

Abstract Dynamic regulation of RNA molecules is critical to the survival and development of cells. Messenger RNAs are transcribed in the nucleus as intron-containing pre-mRNAs and bound by RNA-binding proteins, which control their fate by regulating RNA stability, splicing, polyadenylation, translation, and cellular localization. Most RBPs have distinct mRNA-binding and functional domains; thus, the function of an RBP can be studied independently of RNA-binding by artificially recruiting the RBP to a reporter RNA and then measuring the effect of RBP recruitment on reporter splicing, stability, translational efficiency, or intracellular trafficking. These tethered function assays therefore do not require prior knowledge of the RBP's endogenous RNA targets or its binding sites within these RNAs. Here, we provide an overview of the experimental strategy and the strengths and limitations of common tethering systems. We illustrate specific examples of the application of the assay in elucidating the function of various classes of RBPs. We also discuss how classic tethering assay approaches and insights gained from them have been empowered by more recent technological advances, including efficient genome editing and high-throughput RNA-sequencing.

Keywords Tethered function assays • MS2 coat protein • Lambda N • BoxB • Nucleocytoplasmic transport • Subcellular localization • Translation • RNA stability

T.J. Bos • J.K. Nussbacher • S. Aigner
Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA
e-mail: saigner@ucsd.edu

G.W. Yeo (✉)
Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California, La Jolla, CA, USA

Molecular Engineering Laboratory, A*STAR, Singapore, Singapore

Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
e-mail: geneyeo@ucsd.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_3

1 Introduction

Since the development of tethered function assays in the Wickens lab, where they were first applied to establish the independence of the poly(A) binding protein's RNA-binding activity on its mRNA stabilization activity [1], they have been used to investigate the roles of RBPs in all aspects of RNA metabolism. They have been used to assign specific roles to putative RBPs, dissect the function of individual RBPs within large complexes and separate functional domains of multi-domain RBPs. Because of the ease with which RBPs can be tethered to different sites on the reporter, the assay has also been invaluable in probing the dependence of RBP function on site-specific recruitment to target RNAs. Although tethered function assays have been used predominantly for RBPs, in principle the protein of interest itself does not need to have RNA-binding activity, since the artificial tether accomplishes recruitment. This allows the study of protein complexes that are recruited by RNA-binding proteins. Several well-characterized RNA-binding moieties and the cognate RNA elements they bind have been adapted as the tethers in these assays, and a handful are frequently employed.

Tethered function assays have three practical advantages over alternative methods to investigate RBP function [2, 3]. First, they are more straightforward to set up, carry out and interpret than approaches that rely on genetic manipulation, gene expression knockdowns or overexpression. Tethered function assays allow studies of essential RBP genes, do not require genetically tractable cell types, and circumvent pleiotropic effects of RBP depletion or over-expression on cellular homeostasis that may influence the target gene(s) under study. Second, RBPs can be studied without knowing its endogenous RNA targets. Instead, the effect of RBP binding is investigated with reporter assays that produce robust and straightforward readouts. Specialized reporter constructs enable studying the effect of the RBP of interest RNA in a particular RNA processing step. Third, as RNA metabolism is highly coupled, RBPs that affect a processing step that lies upstream of the one under study often influence the latter. In depletion or over-expression studies, these indirect effect may confound results. However, since these assays depend on artificial reporter constructs, careful experimental controls, independent confirmation of results and follow-up studies are required for meaningful interpretations.

Here, we describe how to design and carry out these assays, provide the strength and limitations of common tethering systems that have been developed to date and illustrate their utility with examples from the literature.

2 Designing and Performing Tethered Function Assays and Interpreting Their Results

Here we describe practical considerations of tethered function assays. For more detailed methods, we refer to two excellent protocol articles for specific application of tethering assays in studying RNA degradation [4] and nonsense mediated RNA decay and translational initiation [5].

2.1 Constructs

A tethered function assay comprises two components. First, a reporter DNA construct is designed which, when expressed in cells, is transcribed into an mRNA that may encode a functional protein (for example GFP, luciferase, or LacZ), depending on the application. The mRNA product also contains an RNA structural element that is recognized by an exogenous RNA-binding moiety. Second, an effector construct encodes the RBP (or RBP domain) fused to this moiety, such that the tagged RBP, when co-expressed, is recruited to the reporter mRNA by binding to the RNA element with strong affinity and specificity. As will be discussed later in this chapter, these RNA structural elements invariably are RNA stem-loop structures—so-called 'hairpins', most commonly from bacteriophage genomic RNAs, which are recognized by the phage coat proteins as their cognate RNA-binding moieties.

Most commonly, following ectopic expression of both constructs in cells, the effect of RBP tethering on target mRNA metabolism is determined by standard molecular biology techniques. For example, the effects of the RBP on reporter mRNA stability or splicing might be evaluated by reverse transcription PCR or northern blot [2, 3], while changes in RNA subcellular localization might be identified by fluorescence in situ hybridization (FISH) [6]. Assays using *in vitro* transcribed mRNA reporters and purified recombinant RBP fusions, both added to cell-free extracts competent for the RNA processing step under study, have also been reported [7, 8]. Lastly, for mRNA localization studies, the transcript of interest has been tagged by inserting recruitment tethers into the 3′ UTR of endogenous genes via homologous recombination in mouse [9]. Figure 3.1 shows an example of a tethered function assay using the bacteriophage MS2 hairpin structure (see below) placed in the 3′ untranslated region (UTR) of a reporter.

2.2 Position and Number of the Tethering Sites

The most crucial factor in the design of these assays is the position of the hairpin in the reporter construct. This decision is usually guided by a hypothesis or some prior knowledge regarding the function of the RBP of interest when bound to natural transcripts. Tethering of RBPs to different regions of the transcript may reveal region-specific RBP functions, with different effects on reporter expression. For example, while tethering of the RBP Staufen (STAU1) to the 3′ UTR of a reporter led to its degradation, 5′ UTR tethering promoted translation [10, 11]. The former observation reflects Staufen's role as an inducer of nonsense-mediated RNA decay (NMD) via recruitment of UPF1, while the latter may be a consequence of Staufen's function as an enhancer of translational initiation of structure-repressed transcripts [10, 12].

Of all regions of a transcript, the 3′ UTR arguably contains the greatest density of regulatory elements. The spatial positioning of an RBP binding site in the 3′ UTR is usually not critical for RBP function, and most tethering studies of RNA turnover, transport and localization have used tethers placed in the 3′ UTR [2, 3].

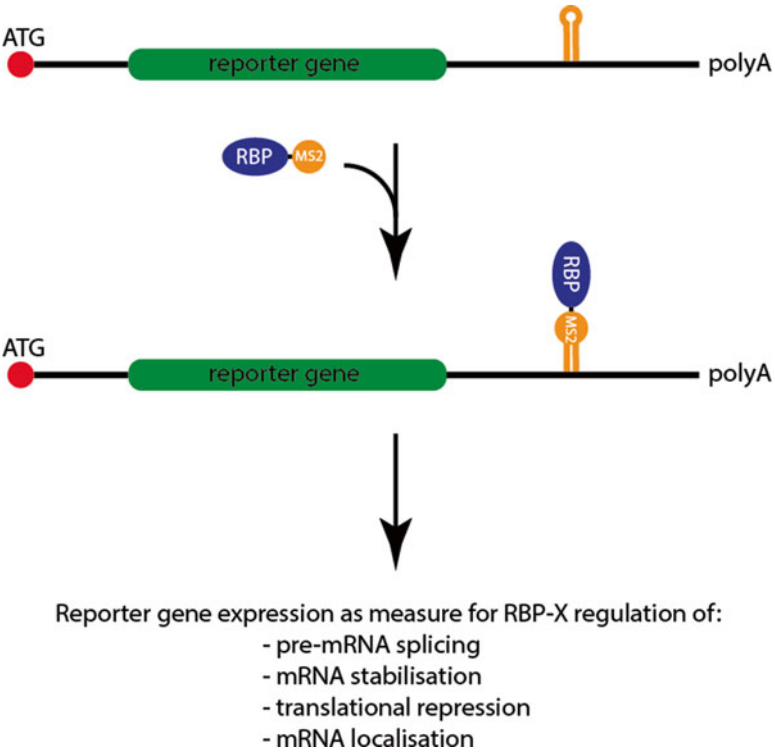


Fig. 3.1 A tethered function assay of an RBP that is tethered to the 3′ UTR. The recruitment of the RBP to the reporter mRNA is artificial via the interaction of the hairpin and hairpin-binding moiety fused to the RBP of interest. The biological effect of binding can be studied by standard molecular biology techniques. In this example, an MS2 tethering system is shown, but alternatives exist

In contrast, an exploration of RBP function in translational initiation will insert the tether in the 5′ UTR [13, 14] or in the intervening sequence between the two coding portions of a bicistronic construct [5, 15]. And as we will see later in this chapter, the functional outcome of splicing factor recruitment to pre-mRNA processing is exquisitely sensitive to the precise binding site relative to the spliced exon; here, the hairpin tether is placed in introns and exons at various distances from the splice junctions.

Although a single hairpin is often sufficient for productive RBP recruitment in cells, the increase in stability of the interaction afforded by cooperative binding of phage coat proteins bound to two hairpins of the MS2 phage (see below) was required for biochemical purification of bound complexes [16]. For RBPs functioning in RNA degradation (including NMD), a commensurate relationship between the number of tethers and the magnitude of the decrease in reporter signal was found [17–20]. In contrast, translational stimulation was reported to increase with

the number of hairpins recruiting the germ cell-specific DAZL protein to the reporter 3′ UTR [21], while this was not observed for the histone mRNA-specific 3′ UTR binding protein, SLBP; here a single tethering site sufficed for maximal stimulation [22]. Thus, there is a consensus that multiple hairpins are generally advantageous and may maximize the number of RBPs recruited, thus increasing signal-to-noise and the likelihood of observing an effect on the reporter. Most studies have employed several (three to six) hairpins in a tandem array, but as many as 24 have been used in single-molecule mRNA localization studies [9]. However, long arrays of hairpins significantly lengthen the target gene region (for example, the 3′ UTR or an intron) of the reporter beyond its natural size; such highly engineered constructs may no longer be recognized and processed by cellular machinery that needs to act on the reporter prior or subsequent to binding of the RBP of interest. For example, an analysis of RBP function in deadenylation-mediated mRNA degradation requires the reporter mRNA to be properly cleaved and adenylated before RBP action. Pilot experiments, using tethered RBPs with known effect on the particular aspect of mRNA metabolism under investigation, in conjunction with reporters containing different hairpin numbers, will identify optimal conditions.

2.3 Limitations, Controls, and Interpretation

Tethered function assays are not suitable for all RBPs. The RNA-binding sites of helicases and nucleases largely overlap with the active sites of these enzymes, so their 'RNA-binding' and 'functional' domains are not readily separable. In some cases, the RBP must bind the mRNA at a specific site or in a specific geometry, or be able to move freely along the RNA substrate or cycle in and out of a complex; here, artificial tethering may fail to properly localize the RBP. RBPs often require additional components to form a functional effector complex and these cofactors may not be expressed in the cell type studied. Even if expressed, they may fail to productively assemble on the reporter, for example if RBP binding to its natural RNA target induces a conformational change in the RBP that is required for their recruitment [23, 24]. And finally, the RNA-binding moiety, the RBP, or the hairpin may be occluded or misfolded in these artificial constructs; these challenges can be addressed in part by exploring different RBP-tag fusion constructs (for example, N- vs. C-terminal fusion and different linker sequences between the RBP and the tag) or tethering systems, and placing the hairpin at different locations on the reporter. Thus, in general, only positive results of tethering assays are meaningful.

Standard negative controls include expression of (1) the untagged RBP along with the reporter, (2) the RNA-binding moiety alone along with the reporter, and (3) of a reporter that either lacks the hairpin or, preferably, contains a binding-incompetent hairpin mutant in the presence of the tagged RBP. These controls ensure that the observed effect on reporter mRNA metabolism is indeed due to RBP

recruitment, rather than due to overexpression of the RBP or the RNA-binding moiety in the cell, or the presence of a hairpin in the reporter *per se*. Tethering hairpins are relatively weak RNA secondary structures which nevertheless may interfere with RNA processing or translation, particularly when placed in the 5′ UTR or the coding sequence (CDS). A careful tethering study will quantify this effect by comparing the activity of the untagged reporter to that of a reporter harboring (a) binding-incompetent mutant hairpin(s), before employing the functional tethering reporter with the same number of functional tethers [14]. For tethering assays at the 5′ UTR, steric hindrance of translational initiation by the presence of a protein, presumably independent of its identity, near the cap has been suggested, as increasing the distance between the cap and the tethering site relieved the observed translational inhibition [25].

Finally, as is the case with all experiments involving ectopic expression of reporter and effector constructs, their relative and absolute expression levels need to be carefully titrated so as to maximize specific reporter signal, while minimizing the risk of artifactual results stemming from their (over-) expression in the cell.

2.4 Follow-Up and Validation

Tethering assays rely on engineered and therefore artificial RNAs and proteins; insights gained from them are therefore only meaningful when confirmed by alternative approaches. As we will see later in this chapter, where we describe examples of tethering assay applications, these specific approaches depend on the biological question; however, they almost always involve investigation of RBP function in the context of its natural RNA targets. If endogenous targets are unknown, cross-linking immunoprecipitation followed by RNA sequencing (CLIP-seq) [26, 27] in relevant cell types might be performed to identify them (see also van Nostrand *et al.*, Chap. 1, this Volume). Knockdown of the RBP of interest, followed by global transcriptome analyses or more specialized phenotypic readouts, will provide clues to the role of the RBP in cellular homeostasis. Finally, biochemical purification of the recruited protein complexes might be used to reveal additional components of the effector complex of the RBP [16].

3 Tethering Systems

Although the bacteriophage MS2 system is the most widely used tethering system [16], several other methods (primarily of bacteriophage origin) are available. In the next section, we will discuss the most commonly used systems in detail, together with their advantages and drawbacks. Table 3.1 presents a summary of these tethering systems along with their characteristics.

Table 3.1 Key characteristics of hairpins and hairpin binding moieties used in tethering assays

Name	Minimal protein size (aa)	Dimer/monomer	Hairpin size (nt)	Dissociation constant (K_d)[a] (Ref.)
MS2	129	Dimer	21	10^{-9}–10^{-8} M [28]
Λ N	22	Monomer	15	10^{-9}–10^{-8} M [29]
PP7	127	Dimer	25	~10^{-9} M [30]
TAT/Tar	17	Monomer	28	10^{-9}–10^{-8} M [31]
IRP	IRP1 889	Monomer	30	10^{-12}–10^{-11} M [32]
	IRP2 964			
U1A	102	Monomer	29	10^{-11}–10^{-10} M [33]
Qβ	129	Dimer	20	10^{-9}–10^{-8} M [34]
GA	129	Dimer	23	10^{-9}–10^{-8} M [35]

aa amino acids, *nt* nucleotides, K_d equilibrium dissociation constant
[a]For the wild-type interaction

3.1 MS2

RNA bacteriophages are small viruses with an icosahedral shell that are capable of infecting bacteria. Their single-stranded RNA genome is roughly 3500 nucleotides (nt) long and encodes structural proteins as well as proteins for viral maturation, replication and the lysis of their bacterial host. These phages were initially isolated from *E. coli*, but are also found in other species of bacteria. The role of the coat protein is twofold: (1) it is the major structural protein of the viral particles and (2) it acts as a translational repressor of the viral replicase. During the late stages of *E. coli* infection by the bacteriophage, translation of the replicase gene is repressed due to binding of the MS2 coat protein to specific sequence elements in the replicase mRNA. This interaction has proven in both its affinity and specificity to be ideally suited for the tethering system.

The popularity of the bacteriophage MS2 (or the closely related R17) tethering system tethering system can largely be credited to its physical and functional characteristics: (1) the MS2 coat protein (MCP) is relatively small (14 kDa, 129 amino acids) thus reducing its potential to interfere with the function of the fused RBP, (2) it binds its 21 nt hairpin (Fig. 3.2a) with high affinity ($K_d = 10^{-9}$–10^{-8} M) and selectivity, limiting potential off-target binding, and (3) the MS2 hairpin-MCP interaction is well-characterized [16, 28, 36]. In addition, mutations in both the MS2 hairpin and the coat protein have been identified that increase the interaction affinity. A single U to C substitution in the loop increases binding affinity by 50-fold over wild type (Fig. 3.2b) [36–39]; this mutant is commonly used in tethering assays. Conversely, a mutant lacking the single-stranded ('bulged') adenosine within the stem essentially abolishes binding activity [36, 40, 41] and is sometimes used as a negative (recruitment-deficient) control in tethering assays [7]. A V29I mutation in the MCP modestly increases binding strength to the hairpin (by ~7.5-fold), albeit with an apparent commensurate decrease in specificity [42].

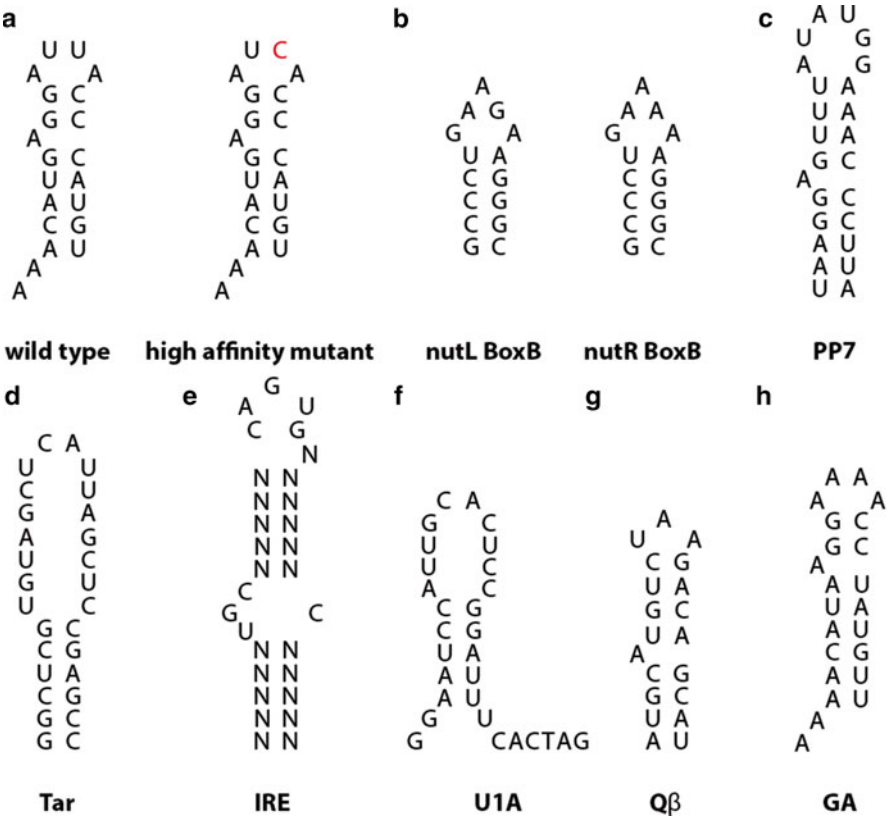


Fig. 3.2 Sequences of all mRNA hairpins discussed in this chapter. (**a**) Wild type (*left*) and the high affinity mutant (*right*) of the MS2 hairpin. (**b**) The 15 nt boxB elements of the λ*N* system. Both versions have been used for tethering and show a similar affinity for the λ*N* protein despite the single nucleotide difference. (**c**) and (**d**) The PP7 and TAR hairpins, respectively. (**e**) A canonical structure of the iron-responsive element (IRE). N denotes base pairs whose nucleotide identities are not critical for IRP binding. (**f**, **g**, **h**) The hairpins of the U1A, Qβ and GA phages, respectively

MCPs bind the hairpin as pre-formed dimers, thus recruiting two copies of the fused RBP of interest to the tethering site. Bound coat protein dimers interact cooperatively with one another when tandem arrays of hairpins are present [43]. These properties maximize RBP occupancy but may prevent tight control of recruitment. At concentrations above ~1 μM, MCP dimers assemble into stable capsid-like structures that are not in equilibrium with soluble dimers and do not bind RNA, thereby decreasing the apparent affinity at high capsid concentrations [40]. Given that even moderately abundant endogenous proteins have intracellular concentrations in the millimolar range [44], it is surprising that capsid formation does not appear to be a problem in tethering assays, where the MCP-RBP fusion protein is usually over-expressed. Mutation screens have

identified MCP mutants that dramatically increase the threshold concentration for capsid formation [42, 45–47], some without affecting the RNA-binding specificity or affinity; however, these are no widely used in tethering assays.

3.2 λN

The bacteriophage λ is the most well-characterized of the lambdoid phages, a family of bacterial DNA viruses. Upon infection, phage promoters are sequentially active, and regulation is mediated by the synthesis of anti-termination proteins. The transcriptional repressor *cro* and the anti-terminator *N* are expressed from the pR and pL promoters, respectively. In the absence of anti-terminator *N*, transcription is stopped at the terminator sequences downstream of the N and cro genes (Fig. 3.3a). To activate the anti-termination function of *N* and enable the expression of the delayed early genes of the bacteriophage, *N* must bind RNA polymerase at the nutL and nutR sites to facilitate read-through at the terminator sequences (Fig. 3.3b).

Both the nutR and nutL regions contain two conserved sequence elements of which only the boxB element forms a 15 nt hairpin to which protein *N* binds (Fig. 3.2b) [48]. Upon binding of *N* to the boxB element of nutR and nutL, additional bacteriophage and host proteins are recruited to achieve expression of the downstream genes. Since its first application for RBP tethering [49], the λ*N*/boxB system has gained popularity due to the extremely small size of the *N* protein (12.2 kDa), which has been suggested to reduce the risk of interfering with the function of the fused RBP [19]. Of the 107 amino acids that constitute the *N* coat protein, the 22 residues of the RNA-binding

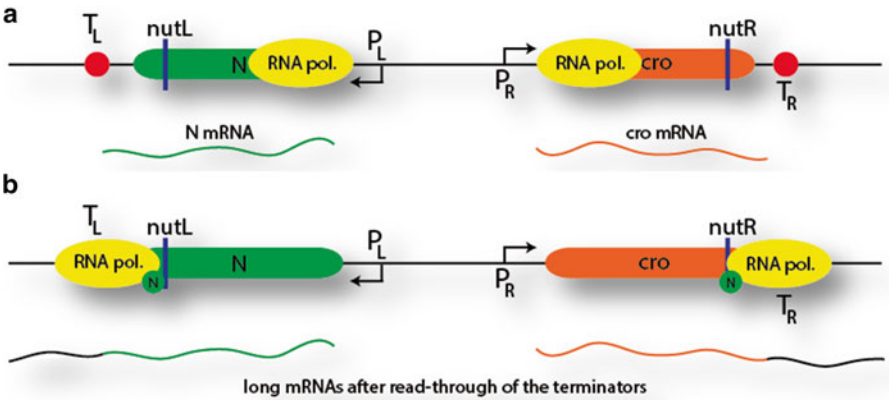


Fig. 3.3 λ*N* anti-terminator function. (**a**) The expression of the immediate early genes N and cro. The RNA polymerase is released from the DNA at both terminator sequences T_L and T_R. (**b**) Read-through of the terminators occurs when the *N*-protein associates with the RNA polymerase at the nutL and nutR sites. This results in the extended transcription into the delayed early genes of bacteriophage λ. P_L and P_R promoter *left* and *right*, *nutL* and *nutR* binding sites for the anti-terminator, T_L and T_R *left* and *right* terminator sequences

domain are crucial for RNA recognition [29]. A synthetic peptide consisting of these amino acids binds the boxB element with affinity and specificity similar to those of its full-length counterpart ($K_d = 10^{-9}$–10^{-8} M) [50, 51]. Additionally, a triple mutant of the peptide (M1G D2N Q4R) with a ~70-fold increased affinity at physiological monovalent cation concentration has been designed [43] but it is unclear to what extent specificity is maintained. In contrast to the MCP, the λ*N* protein binds its RNA element in a 1:1 stoichiometry [52], and therefore non-cooperatively, which may be advantageous for experiments requiring tight control of target occupancy. Although direct comparisons have not been done, it appears that the λ*N*/boxB system performs very similar to the MCP/MS2 system.

3.3 PP7

PP7 is a single-stranded RNA bacteriophage of *Pseudomonas aeroginosa* and is a distant relative of the MS2 and Qβ bacteriophages (see below). Similar to MS2, the PP7 coat protein is a translational repressor of the viral replicase gene. The PP7 coat protein binds its hairpin (Fig. 3.2c) with a K_d of ~10^{-9} M [30], in line with other phage-derived tethering systems; however, the sequence homologies between the PP7 and MS2 coat proteins and their RNA hairpins are very limited [53]. Although the PP7 system has been used on its own in standard tethering assays [54], its strength lies in its compatibility with the MS2 system: the PP7 and MS2 coat proteins discriminate against the respective non-cognate hairpin with ~1000-fold specificity [40, 53]. This orthogonality of the PP7 and MS2 systems has allowed real-time imaging of allele-specific transcription dynamics in yeast. In a diploid yeast strain, one allele of *MDN1* was tagged with 24 MS2 hairpins in the 3′ UTR, the other with 24 PP7 hairpins. Transcripts produced from the two loci were distinguished using MCP and PP7 fusions of two different fluorescent proteins [55]. Similarly, fluorescence complementation at the site of recruitment of two halves of a single fluorescent protein, tagged with MS2 and PP7 coat protein moieties, allowed for reduction in background fluorescence of unbound probe in live-cell RNA localization studies [56].

3.4 Iron Responsive Protein (IRP)

Iron is essential for life. It is primarily involved in the synthesis of heme and hemoglobin in erythroid cells of most vertebrates [57], enabling transport of oxygen throughout the body. Both proteins are involved in the uptake (transferrin receptor) and storage (ferritin) of iron and are translated from transcripts containing one or more iron responsive elements (IREs) in either of their UTRs. These elements consist of roughly 30 nucleotides that are highly conserved in vertebrates, some insects and many bacteria, and form stem-loops. A consensus sequence and structure for

IREs has emerged and comprises a 6-nt loop motif (5′-CAGWGH-3′, where W = A or U and H = A, C or U) that sits atop a 5-nt loop-proximal stem, which itself is separated from the distal stem by an asymmetrical bulge that contains an unpaired C (Fig. 3.2e) [58]. In iron-starved cells, IREs are bound by proteins IRP1 and IRP2, which results in translational repression for IREs located in the 5′ UTR or RNA stabilization for IREs located in the 3′ UTR. The mammalian transferrin receptor 1 (Tfr1) mRNA contains five such IREs in its 3′ UTR, while both the ferritin heavy (H) and light (L) chain transcripts each contain a single IRE in their 5′ UTRs. Iron starvation thus results in (1) the stabilization of the otherwise unstable transferrin receptor mRNA and (2) the translational inhibition of both chains of the ferritin protein messages [59]. This causes increased uptake of iron and also minimizes the sequestration of iron into ferritin (for a review, see [60]). Upon repletion of iron, the transferrin receptor mRNA is degraded and iron is stored by ferritin (Fig. 3.4).

The iron response protein (IRP) tethering system was first used by the Hentze lab [49]. Similarly to PP7, it is not widely used but its compatibility with other tethering systems has enabled recruitment of two different RBPs to two different locations within the same reporter. Gehring *et al.* [19] used such a dual reporter to show that a reduction in reporter mRNA levels, observed when λ*N*-fused UPF3B was tethered to its 3′ UTR via five boxB sites, depended on translation – an observation that would support UPF3Bb—triggered NMD of the message. To this end, they incorporated a single IRE in the 5′ UTR, which allowed recruitment of *endogenous*

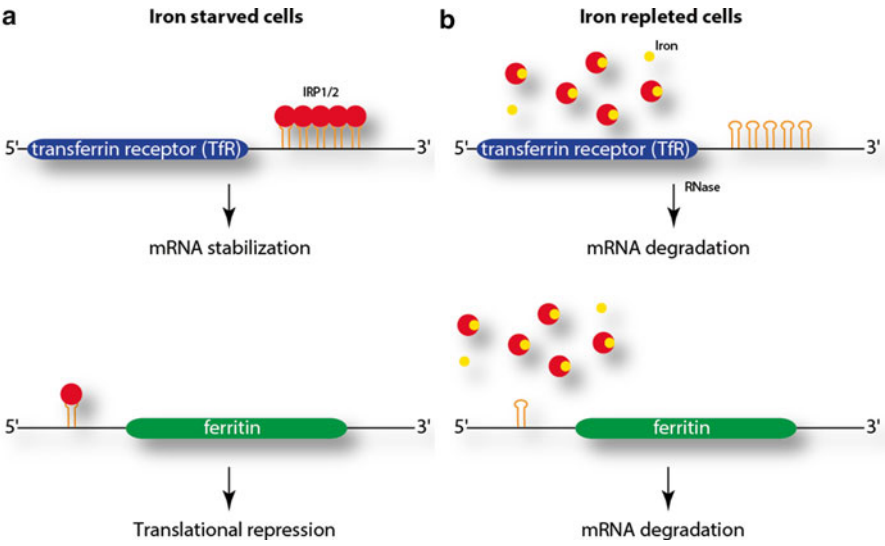


Fig. 3.4 The iron metabolism in humans. (**a**) In iron starved cells, the transferrin receptor mRNA is bound by IRP 1 and/or 2 proteins, resulting in the stabilization of the mRNA and thus expression of the receptor. The ferritin (responsible for storing iron) mRNA is also bound by IRP1 and/or 2 and is subsequently repressed by translational inhibition. (**b**) In iron depleted cells, the situation is reversed: iron taken up by the cell is stored in ferritin stores, while the transferrin receptor expression is inhibited by mRNA degradation

IRPs. Inhibition of translation by iron depletion in cultured cells harboring the reporter restored transcript levels, as expected for an NMD target. In addition to its suitability in such duplex (or multiplex) tethering assays, the IRP system's advantage may lie in the higher affinity of the IRP proteins for its hairpin ($K_d = 10^{-12}$–10^{-11} M) compared to phage coat protein systems. Although not tested, this high affinity may aid in biochemical purification of co-recruited proteins. On the other hand, the IRPs are large; fusing a bulky protein to an RBP might alter its conformation, or hinder RNA target or protein cofactor binding. A total of 12 mammalian mRNAs have been reported to contain IREs, most coding for proteins with known roles in iron metabolism [59], and binding of the to these mRNAs may alter cellular gene expression and thereby confound results.

3.5 Qβ, GA, Tat/TAR, and U1A

Here, we briefly draw attention to four alternative tethering systems that have not been extensively utilized in the field; they possess unique and useful features that may prove advantageous for certain experimental systems or when multiplexing of tethering systems is required.

Qβ and GA. These are derived from coat proteins of the eponymous bacteriophages which, although only distantly related to MS2, show extensive structural similarity as well as common features for recognition of their cognate hairpins [61, 62, 63]. Two mutants of the Qβ coat protein, D91N and Q65H, have shown a higher affinity for its hairpin (Fig. 3.2g) than wild-type but they also bind the MS2 hairpin more efficiently [64]. On the other hand, the MCP has a 100-fold greater affinity for its own hairpin than for the GA-associated hairpin (Fig. 3.2h). In contrast, the GA protein binds its own and the MS2 hairpin with the same affinity [65]. Given the high degree of similarity between the MS2, Qβ and GA systems, the use of MS2 is preferred because it is better characterized.

Tat/TAR. One of the few non-bacteriophage tethering systems is derived from the bovine immunodeficiency virus (BIV). Only 17 amino acids from the BIV Tat (transactivator of transcription) protein are necessary to bind the 28-base BIV TAR (trans activator response) element (Fig. 3.2d) with high affinity ($K_d = 10^{-9}$–10^{-8} M) [66]. Wakiyama and colleagues [67] provided proof-of-principle for the utility of this system by showing that tethering to the 3′ UTR of Tat-fused TNRC6B (a member of the GW182 family of proteins functioning in microRNA-mediated gene repression) via four TAR elements effected a magnitude of reduction in reporter activity similar to that seen with λN-fused TNRC6B and a 5x boxB reporter. No other groups have since used it for tethering proteins to a reporter, likely due to the widespread use of better characterized tethering systems.

U1A. The N-terminal 100 amino acids, containing one of two RNA recognition motifs (RRM), of the human U1 small nuclear ribonucleoprotein (snRNP)-specific protein U1A (also known as SNRPA), binds with high specificity and affinity to a

~25 nt hairpin structure (Fig. 3.2f) in U1 snRNA, called U1A hpII (Fig. 3.2f). Its 5′-AUUGCAC-3′ single-stranded loop sequence is critical for binding [68]. Because yeast U1 snRNA lacks this hairpin [69], the RNA-protein interaction has mainly been employed in yeast for tracking subcellular localization of RNA transcripts and studying mRNA processing [70–73].

4 Applications of Tethered Function Assays

In what follows, we highlight select publications that illustrate the versatility of tethered function assays, as they have been used in virtually all areas of RNA processing and RBP research. As we have seen, the development of tethered fluorescent proteins probes for tracking RNA subcellular localization is an active area of investigation, and we provide additional examples this extension of canonical tethered function assays.

4.1 mRNA Stability

Poly(A) binding protein: functional analysis of an essential gene. A tethering assay was first developed by Coller *et al.* [1] to study the role of the poly(A)-binding protein (Pab1p) in yeast. Previous studies had shown that *PAB1* is an essential gene and that lethality of a *pab1* deletion could be supressed by mutations in in mRNA decay-related genes, suggesting that Pab1p protects mRNAs from degradation. However, since *PAB1* is essential, genetic studies could not be used to investigate whether mRNA stabilization simply required Pab1p's presence on mRNAs, or whether this function was dependent on Pab1p's interaction with the poly(A) tail. With yeast strains expressing non-essential reporters under transcriptional pulse-chase conditions, Coller *et al.* [1] showed that while Pap1p, when fused to an MCP dimer fusion and recruited to reporters via two MS2 hairpins, stabilized both a deadenylated reporter mRNA and one that cannot be adenylated (Fig. 3.5), it did not affect the reporter deadenylation rates [1]. Thus, Pab1p stabilized mRNAs independent of their poly(A) tail, suggesting a very simple model: the poly(A) tail serves as a binding site for Pab1p molecules; deadenylation removes these Pab1p sites and upon departure of the last Pab1p molecule, RNA degradation is triggered. Using a reporter with a stem-loop structure in the 5′ UTR that rendered it translation-incompetent, the authors also showed that mRNA stabilization by Pab1p requires some aspect of translation. It is now well established that the mRNA 3′ end processing machinery and translational initiation complexes are physically and functionally and linked [74].

Exon junction complex proteins: linking nonsense-mediated RNA decay to splicing. One of the most important mRNA quality control checkpoints in the cell is the nonsense-mediated mRNA decay pathway (NMD). In a manner that is dependent on prior splicing, exon-exon junctions are marked by deposition of the exon junc-

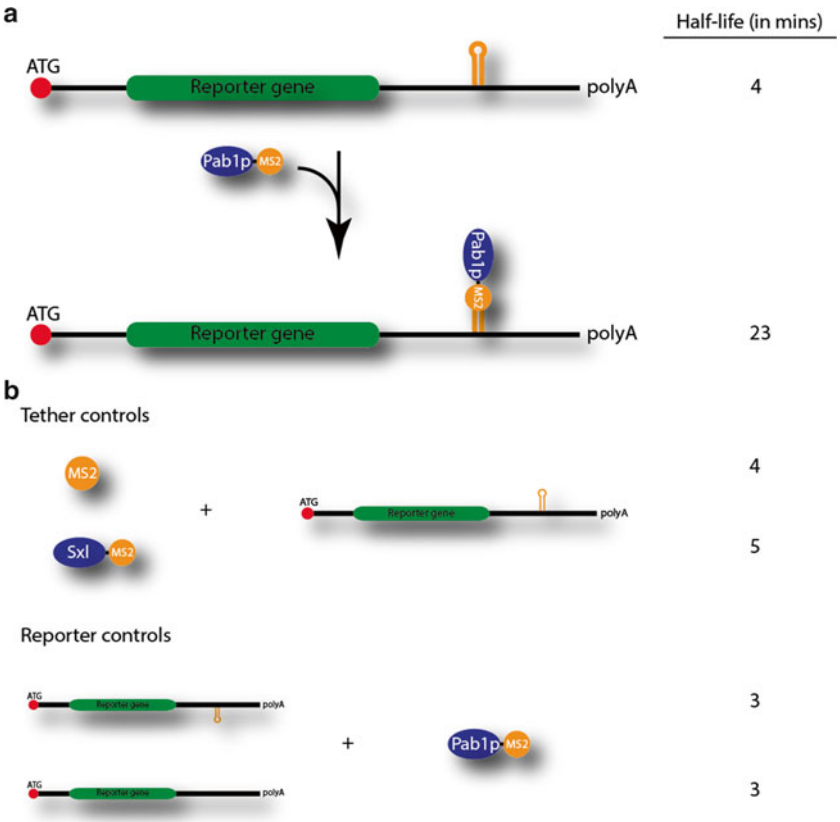


Fig. 3.5 Overview of the mRNA stability experiments performed by Coller *et al*. [1] (**a**) By tethering Pab1p to an unstable reporter gene, an increase in stability was observed. (**b**) This example is a good illustration of how controls should be included in a tethering assay. In their study, Coller *et al*. used both tether controls and reporter controls. The tether controls evaluate the effect of the tether on the reporter; this is determined by tethering either an MS2 alone or an MS2 protein fused to a protein of similar size as Pab1p (Sxl in this case). The reporter controls determine whether the observed increase in mRNA stability is the result of the tethering of the protein. This can be achieved by using a reporter mRNA without a tether hairpin and/or a reporter with the hairpin in the antisense orientation. Adapted from [1]

tion complexes (EJCs) [75–77], which allow cells to distinguish legitimate stop codons from premature termination codons (PTCs) (Fig. 3.6a). A termination codon located more than 55 nt upstream of the last exon-exon junction is sensed as premature, subjecting the mRNA to NMD. Lykke-Andersen *et al.* [18] used a human β-globin mRNA tethering reporter to validate candidate human homologs of three yeast proteins, termed Up frameshift (Upf) proteins 1–3, which had been identified through genetic screens as suppressors of a frameshift mutation [78]. MS2-based tethering of the candidates to the reporter 3′ UTR circumvented the requirement for prior splicing. Using this system, the authors then asked which of

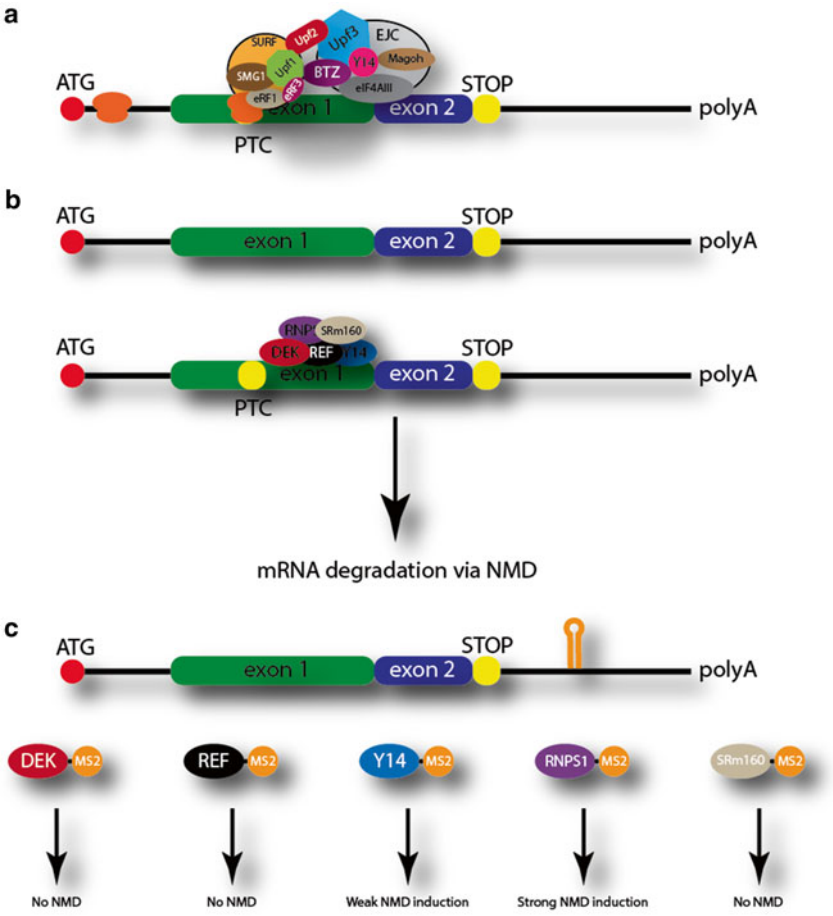


Fig. 3.6 The tethered function assay used for studying NMD. (**a**) Schematic overview of the NMD mechanism. The EJC complex (consisting of eIF4AIII, Y14, MAGOH and BTZ and Upf3) recognizes the exon-exon junction between exon 1 and exon 2. Next, the SURF complex (consisting of PIKK (not shown), SMG1, Upf1, eRF1 and eRF3) is recruited to the premature termination codon (PTC) and initiates transcript degradation. SMG5, SMG6 and SMG7 are not shown in this figure. (**b**) The top construct of the panel shows the wild type situation: 2 exons with a stop codon downstream resulting in a stable mRNA. The bottom construct illustrates what happens if a PTC is present. Upstream of the exon-exon junction, at least 5 proteins bind the mRNA as a complex and trigger NMD. (**c**) Tethering each individual protein downstream of a wild type stop codon, to the 3′ UTR mimics NMD and therefore, mRNA stabilization by each individual protein of the complex can be studied. Protein RNPS1 and to a lesser extent Y14 elicit NMD, while the other three proteins (DEK, REF and SRm160) do not alter mRNA stability. From [79]

seven proteins that had been reported as members of a multiprotein post-splicing complex were required for linking NMD to splicing via binding to Upf(s) (Fig. 3.6b) [79]. The RBPs RNPS1 and, to a lesser extent, Y14 (RBM8A) triggered NMD when tethered to the reporter 3′ UTR (Fig. 3.6c) and both interacted with Upf proteins in

co-IP assays. It is now known that Y14 is part of the heterotrimeric core EJC, while RNPS1 is a member of the EJC-peripheral ASAP effector complex, which links the EJC to mRNA quality control pathways and the translational machinery [80].

YTHDF2: Understanding the mechanism of RNA destabilization by an RNA modification reader protein. Adenosine N6 methylation (m^6A) has recently been identified as an additional layer of posttranscriptional gene regulation. Members of the YTH domain family (YTHDF) of RNA-binding proteins recognize m^6A in RNAs and have been identified as m^6A binding proteins by affinity chromatography. Wang et al. [81] used λ*N*/boxB-based tethering assays to investigate the outcome of forced recruitment of a YTHDF2 construct lacking its C-terminal RNA-binding domain, to a reporter mRNA containing five boxB hairpins in its 3′ UTR. They observed a reduced steady-state level and shorter poly(A) tail of the YTHDF2-bound reporter. These findings, along with results from knockdown and immunocytochemistry experiments led them to propose that YTHDF2 targets m^6A-containing mRNAs to sites of RNA decay, such as processing bodies.

4.2 Translation

eIF4F complex: function of core translation initiation factors. Translation initiation is the rate-limiting step in translation and thus most translational regulation occurs at this stage. In cap-dependent translation, the cap-binding complex eIF4F, consisting of eIF4E, -G and -A, recruits a preassembled 43S ribosomal preinitiation complex (PIC), which then scans the mRNA 5′ end for a start codon. It is now known that eIF4E recognizes the cap and binds to eIF4G, which in turn recruits the PIC via interacting with eIF3, a component of the PIC [74]. Tethered function assays proved invaluable in dissecting the roles of the eIF4F complex subunits. In a study that pioneered the use of both λN/boxB and IRP/IRE tethering systems, the Hentze lab showed that forcible recruitment, to the intercistronic space of a bicistronic reporter, of an eIF4GI mutant lacking the eIF4E binding domain supported translation of the downstream cistron [49]. Similar results were obtained with an eIF4E mutant lacking the cap binding domain, but not with eIF4E lacking the eIF4G binding domain or with full-length eIF4A [49]. These studies showed that eIF4G and eIF4E are sufficient for translational initiation in the absence of cap binding. The failure of tethered eIF4A to promote translation was attributed to the fact that this helicase might require flexible association with the RNA, which tethering—trivially—prevents. As described above, this finding underscores the limitations of tethered function assays for helicases.

AGO2: mechanism of microRNA-mediated gene silencing. microRNAs (miRs) and endogenous siRNAs are important regulators of transcription and translation. Both classes are found in RNA-protein complexes at their target mRNAs with GW182 and Argonaute (Ago) proteins at their core [82], which recruit additional factors to form the RNA-induced silencing complex (RISC) [83]. In a series of tethered function assays using a λN/boxB-based luciferase reporter, the Filipowicz and Izzauralde

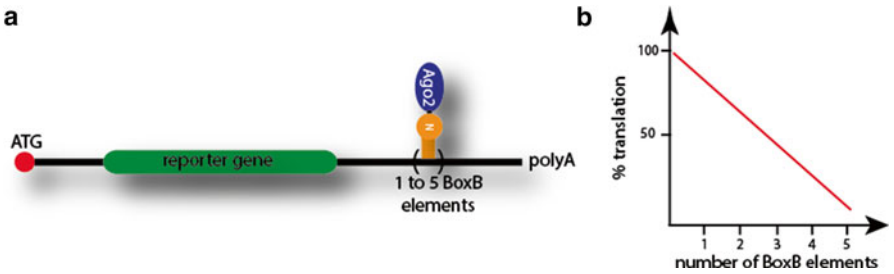


Fig. 3.7 Studying mRNA translational inhibition by Ago protein via a tethering assay. (**a**) Downstream of a reporter gene, 1–5 BoxB elements are inserted. By fusing the 22 N-terminal amino acids of the λ N protein to the Ago2 protein, Ago2 can be tethered to the reporter and inhibit translation. (**b**) This graph illustrates that increasing the number of boxB elements in the 3′ UTR of an mRNA proportionally decreases translation. From [17]

labs elucidated the roles of RISC and accessory components in gene silencing and shed light on its molecular mechanism [17, 84, 85] (Fig. 3.7a). Tethering of AGO1, AGO2, AGO4 and GW182—but not HIWI (PIWIL1), a Paz and PIWI domains protein like Ago proteins—to the 3′ UTR of a luciferase reporter via the λN/boxB system reduced luciferase activity in a manner dependent on the number of boxB elements (Fig. 3.7b), but not their position within the 3′ UTR. Upon siRNA-mediated depletion of GW182, but not of AGO2, repression of an AGO1-tethered reporter was partially relieved. These simple assays provided fundamental insights into mechanism of miRISC-mediated silencing and provided the groundwork for many later in findings from global analyses, such as the additive effect of multiple miRs bound to target mRNAs, the functional redundancy of AGO1 and AGO2, and the importance of GW182, which is now known to interact with PABP and deadenylase complexes. They also provided the first insights into the contribution of both mRNA degradation and translational inhibition to miR-mediated gene silencing.

TYF: molecular mechanism of PER translational stimulation. Tethered function assays have played a central role in appreciating posttranscriptional mechanisms in controlling of circadian rhythms. The product of the *Drosophila* gene 24 (TYF) was identified in a mutation screen to be necessary for robust behavioral rhythms in pacemaker neurons. Lim et al. [8] observed that the core circadian clock protein PERIOD (PER) depended on the expression of TYF, which had no known functional domains. To understand if TYF acts transcriptionally or post-transcriptionally, they employed tethering assays in transfected *Drosophila* cells and, notably, in translation-competent *Drosophila* cell-free extracts. Here, a recombinant MCP-tagged C-terminal fragment of TYF robustly upregulated translation of an *in vitro* transcribed luciferase mRNA reporter harboring six MS2 hairpins in its 3′ UTR. Co-IP of PABP and eIF4E with TYF and polysome profiling studies further corroborated a role for TYF in promoting translation [86]. Co-IP and mass-spectrometry studies further identified ATAXIN-2 (ATX2) as a TYF interactor; tethered ATX2 stimulated reporter activity similar to TYF. These results revealed a

central role for ATX2, an RNA-binding protein whose human homolog is implicated in neurodegenerative disease, in controlling circadian timing at the posttranscriptional level (see also Benegiamo *et al*., Chap. 5, this Volume).

4.3 Pre-mRNA Splicing

SR proteins: functional dissection of RBP domains in cell-free extracts. Graveley and Maniatis [7] were the first to report the use of tethering assays to study splicing and used them to dissect the roles of individual RBP domains. Serine/arginine (SR) proteins are essential components of the intronic splicing machinery. They are bipartite RBPs with one or two RRMs, critical for RNA-binding [87, 88], and one arginine/serine-rich (RS) domain necessary for SR protein function [89–93]. They regulate constitutive splicing by interacting with components of the basic splicing machinery but are also involved in the regulation of alternative splicing events. Splicing enhancers located downstream of the regulated intron can be bound by SR proteins, which enhances splicing of the upstream intron [87, 94]. If placed downstream of an intron, SR binding sites can function as splicing enhancers by recruiting basic components of the splicing machinery. In order to test whether the 'general' and 'regulatory' functions of SR proteins can be uncoupled, Graveley and Maniatis [7] constructed three *in vitro* transcribed mRNA reporters consisting of enhancer-dependent introns with a single MS2 hairpin replacing the downstream splicing enhancer (Fig. 3.8a). Using recombinant proteins consisting of the RS (Arg/Ser-rich) domains of the SR proteins SF2/AF (SFRS1), SC35 (SRSF2), and 9G8 (SRSF7) and the MCP, alternative splicing was observed in HeLa cell nuclear extracts (which contain endogenous SR proteins), indicating that the splicing enhancer function of RS domains is separable from their RNA-binding domains. However, these hybrid proteins were not able to substitute for endogenous SR proteins; in extracts lacking the latter, purified SR proteins were required to observe splicing. The authors thereby showed that the functions of SR proteins in the basic splicing reaction are separable from their role as splicing enhancers [7].

TRA2: investigation of context-dependence of splicing factor binding. In addition to their role as splicing enhancers, certain SR proteins and related factors can act as splicing repressors. For example, the SR-related protein TRA2 binds an intron (termed M1) within its own pre-mRNA and promotes M1 retention [95]. In order to investigate the requirements for splicing repression by TRA2 in cell-free extracts, Shen and Mattox [96] tethered TRA2 to two different locations within an M1 intron reporter and found that both caused intron retention, while exonic positioning downstream of the 3′ splice site supported splicing activation. TRA2 domain analyses using these reporters showed that the C-terminal RS domain (RS2) was sufficient to cause activation (in exonic placement), while repression (in intronic placement) required an intact RRM. These results indicated that TRA2's functions in splicing are context-dependent and likely mediated by recruitment of distinct effector complexes by separable regions of TRA2.

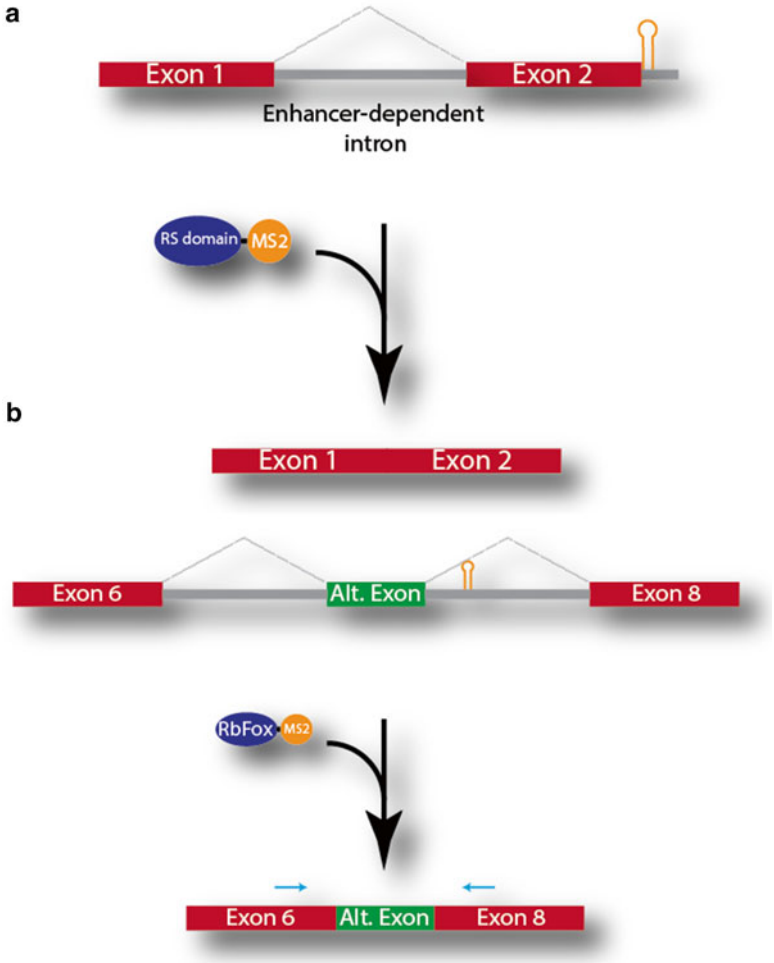


Fig. 3.8 Overview of two different tethering approaches for studying splicing events. (**a**) The reporter construct contains two exons flanking an enhancer-dependent intron. Downstream of exon 2, an MS2 hairpin replaces the endogenous enhancer sequence. The RS domains of three different SR proteins are tethered to the reporter, resulting in the splicing of the intron. From [7]. (**b**) The *SMN2* reporter construct used by Sun *et al.* Exons 6 and 8 of the *SMN2* locus are flanking the alternative exon 7 (the alternative exon in the figure). The hairpin is located in the intron downstream of the alternative exon. Tethering RBFOX proteins to the reporter resulted in the retention of the alternative exon. From [97]

RBFOX1: identification RBP-recruited effector complexes. Sun *et al.* [97] investigated the mechanism of RBFOX1-mediated splicing activation and repression by tethering this alternative splicing factor to different sites on a reporter minigene. Members of the RBFOX family of proteins control exon inclusion and exclusion in a position-dependent fashion: they promote exon inclusion when bound downstream of an alternative exon, and exon exclusion when bound upstream [98, 99]. In order to

understand the molecular mechanism underlying this context-dependent regulation, Sun *et al.* [97] tethered domains of the RBFOX1 protein to a single MS2 hairpin located in introns downstream and upstream of alternative exon 7 of the *SMN2* gene, which is devoid of natural RBFOX binding sites in this region (Fig. 3.8b). While the Ala/Tyr/Gly-rich C-terminal domain of RBFOX1 was sufficient for exon inclusion in the downstream configuration, both the C-terminal domain and the central RNA RRM domain were required for exon skipping in the upstream configuration. Co-immunoprecipitation (co-IP) studies using the C-terminal domain as bait identified several proteins, including the RBP hnRNP H1 and the signalling protein TFG. siRNA-mediated depletion of either protein reduced exon inclusion and exclusion of endogenous RBFOX targets, confirming their role in RBFOX-mediated control of alternative splicing. This study highlights the utility of tethering assays in shedding light on the identities of effector complexes recruited by RBPs.

4.4 RNA Transport and Localization

She2p: control of ASH1 *mRNA trafficking.* One of the best examples illustrating the use of tethering systems in the study of mRNA localization is the yeast *ASH1* mRNA, whose CDS and 3′ UTR contain a total of four *cis*-acting localization elements. During late anaphase, *ASH1* mRNA is localized to the bud tip and sorted to the nucleus of the daughter cell, where its protein product inhibits mating type switching [100, 101]. Genetic studies identified five genes, *SHE1-5*, essential for Ash1p sorting, three of which were known or hypothesized to encode proteins involved in organization of the actin cytoskeleton (*SHE4* and *SHE5/BNI1*) or encode a type V myosin (*SHE1/MYO4*). Long et al. [102] investigated the role of the remaining two She proteins. Tethering She3p in a *she2* deletion strain was sufficient to induce localization to the bud tip, as judged by fluorescence *in situ* hybridization (FISH). Since electrophoretic gel mobility shift assays showed that She2p has binding activity to one of the localization elements, and She2p and She3p interacted in a yeast two-hybrid assay, the authors to proposed that She2p directly binds to the *ASH1* transcript's localization elements factors and associates with a She3p/Myo4p complex to transport *ASH1* mRNA along actin filaments. A schematic overview of this system can be found in Fig. 3.9a, b.

ASH1 *and β-actin mRNAs: monitoring cytoplasmic transport of RNAs.* An adaptation that naturally follows from tethered function assays is the use of forcibly recruited fluorescent proteins to visualize the transport and identify the subcellular localization of tagged RNAs (Fig. 3.9c). In contrast to FISH, this approach allows real time (live-cell) imaging of RNAs and facile detection of their co-localization with proteins. FISH requires fixed (dead) cells and it is often difficult to identify conditions under which both probe hybridization to the RNA and antibody binding to (a) protein(s) of interest are efficient and specific. *ASH1* mRNA was the first RNA to be visualized [103]: GFP fused to a high-affinity mutant of the MCP was

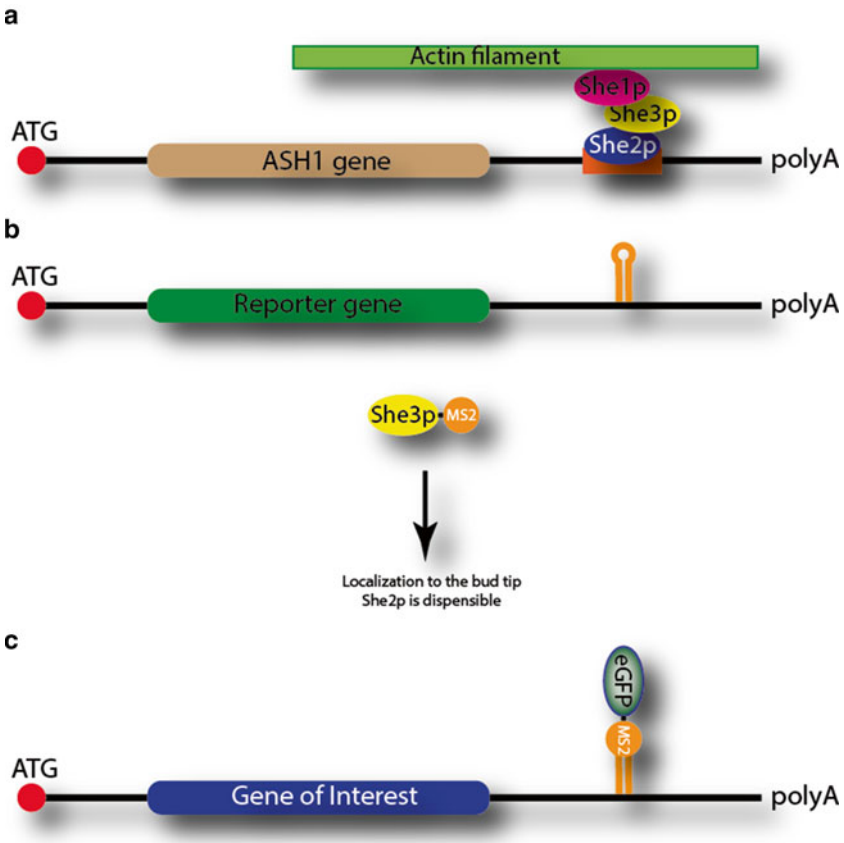


Fig. 3.9 mRNA localization studies by tethered function assay. (**a**) The *cis*-acting element straddling the CDS and 3′ UTR of the *ASH1* gene is bound by She2p, which recruits She3p and She1p to the complex. She1p binds to the actin filament and localizes the mRNA to the bud tip. (**b**) The experiment performed to demonstrate that She2p is the natural tether to the ASH1 mRNA. When She3p is tethered to the reporter, it causes localization of the mRNA in the absence of She2p. (**c**) MCP-GFP fusions can be used to study the localization of an MS2-tagged mRNA of interest in the cell. Of course, other tethering systems can be used for this approach. From [103]

used to track *ASH1* mRNA tagged with six repeats of the MS2 hairpin. The chimeras were expressed from plasmids in yeast strains deleted for either one of the five *SHE* genes, which allowed the dissection of the roles of the She proteins in formation of the *ASH1* mRNA transport complex, its spatial dynamics within the cell, and its anchoring at the bud tip upon arrival. Advances in fluorescence imaging and image analysis have since allowed the visualization of single RNA molecules, perhaps culminating in the ability to track individual transcripts expressed from their endogenous locus in mammalian cells. Lionnet et al. [9] isolated primary embryonic hippocampal neurons from a mouse strain engineered to express β-actin mRNA

tagged with 24 MS2 hairpins from both alleles. Expression of a fusion protein consisting of yellow fluorescent protein (YFP), a nuclear localization signal, and the MCP from a lentiviral vector allowed to follow the motions of single β-actin containing particles. The use of nuclear localized YFP coat protein fusion protein and the large number of MS2 hairpins were required to overcome the fluorescence background from free (non-RNA bound) YFP fusion protein in the cytoplasm. Certainly, a future challenge will be the application of this approach to less abundant transcripts, and the study of mRNA transport dynamics in living organisms.

Monitoring nucleocytoplasmic transport of RNAs. A similar approach allows monitoring the nucleocytoplasmic transport of single mRNP complexes. Mor *et al.* [104] generated inducible gene constructs encoding fusions between a green or blue fluorescent protein and portions of the human dystrophin gene. The constructs also specified 24 MS2 hairpins in the 3′ UTR of the chimeric transcripts. Upon stable expression in mammalian cells, transcripts were detected by transient expression of a plasmid encoding an MCP fused to a red fluorescent protein. In this manner, both the transcript and its protein product could be detected simultaneously and in real time. These experiments showed that mRNP transport to the nuclear pore complex occurs on the minute time scale, consistent with passive one-dimensional diffusion along chromatin channels. In contrast, export is rapid (within 0.5 s), indicating active transport, and is not rate-limiting. This study has enabled significant insights into fundamental aspects of nucleocytoplasmic transport of a model mRNA, which can now be extended to the investigation of the function of endogenous genetic elements, such as sequence motifs in UTRs and introns, in this process.

5 Conclusions and Outlook

Tethered function assays are conceptually simple experiments for rapid functional determination of the effect of a specific RBP on the metabolism of a reporter mRNA. As we have seen, they have been instrumental in all areas of RNA research, where they continue to be employed for the functional dissection of RBP domains and the protein complexes they recruit, and for understanding how the function of an RBP depends on where in a target RNA it is recruited. Of course, validation of tethering assay results by other means is always necessary since the method relies on artificial reporters, which are often ectopically over-expressed from minimal constructs lacking gene features deemed non-essential for the readout of interest. This reductionist approach has nevertheless been useful because it dissects RNA processing events and thus allows the study of the impact of a given RBP on a single event in isolation. In reality, however, RNA processing events are highly interdependent, and so the absence of an RBP effect in a tethering assay for a given process does not necessarily exclude a role for that RBP in this process. For example, the deposition of EJCs, which requires splicing, stimulates translation [105], so the intronless reporters typically used for interrogating RBP effects on translation will not identify RBPs that may affect translation via modulating EJC function. Tagging

of genes at their endogenous genomic loci at scale, now facilitated by CRIPSR/Cas9 genome editing technology, will enable tethering assays that are sensitive to such 'reach-through' effects of RBPs, while at the same time approximating expression levels and processing pathways of natural mRNAs.

More recently, the standard tethering method has been adapted for live-cell tracking of RNAs; although other systems are being developed [106], it is the only approach widely used for real-time imaging of RNAs at the single-molecule level. Here, a superior next-generation approach would have to combine the sensitivity of detection afforded by tandem arrays of hairpins, while simultaneously overcoming the need for such tags in the first place. A related application in which tethering tags may feature very prominently in the future—again facilitated by CRISPR/Cas9—is their use as affinity tags for identification of proteins and nucleic acids in complex cellular RNP structures. Many complex macromolecular assemblies in the cell—including nuclear bodies such as Cajal bodies, speckles and paraspeckles [107]—harbor hallmark RNA molecules that may be harnessed, when tagged with boxB or MS2 hairpin tethers, as baits for the biochemical isolation of interacting protein components, RNAs, and chromatin domains, via affinity purification on λ*N* peptide or MCP columns. An analogous approach termed RNA affinity purification (RAP) used immobilized oligonucleotides that hybridize to Malat1, a non-coding RNA specific to nuclear speckles, and identified RNAs and chromatin associated with these bodies [108, 109]. A tethering method in place of an antisense approach may allow purification under more physiological or more stringent conditions, or at higher yields.

As more and more RBPs are appreciated as playing central roles in disease-related cellular processes [110], interest in elucidating RBP function and targets is increasing accordingly (see also Hattori *et al.*, Chap. 7; Bejar, Chap. 9; and Fan and Leung, Chap. 11; all this Volume). Tethering assays will facilitate identification of these functions, while CLIP-seq approaches [26, 27] will provide the means to elucidate RBP mRNA targets. The combination of both approaches will help uncover if and how the lifecycles of specific RNAs are perturbed when their bound RBPs carry disease-relevant mutations. Clearly, such an approach would require parallel functional interrogation of these RBPs with a battery of tethering reporters—both published and newly developed—that inform on different aspects of RNA metabolism. The ability to conduct tethering assays at scale will also elucidate the effect of specific combinations of RBPs, particularly those with known opposing roles in a given process. Such multiplex assays are already possible because several tethering systems, for example the λ*N*/boxB and MCP/MS2 systems, are orthogonal. It is clear that tethering assays will remain at center stage in RBP research, empowered by genome editing to create physiologically relevant reporters, and by RNA-seq methods to validate their findings.

Acknowledgements The authors would like to thank members of the Yeo laboratory for critical reading of the manuscript. This work was partially supported by grants from the National Institutes of Health (HG007005, HG004659 and NS075449) to G.W.Y. G.W.Y. is an Alfred P. Sloan Research Fellow. J.K.N. was supported by the NCI training grant T32CA067754. T.J.B. is a Hoover Brussels Fellow of the Belgian American Education Foundation (BAEF).

References

1. Coller JM, Gray NK, Wickens MP (1998) mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev 12(20):3226–3235
2. Coller J, Wickens M (2002) Tethered function assays using 3′ untranslated regions. Methods 26(2):142–150
3. Coller J, Wickens M (2007) Tethered function assays: an adaptable approach to study RNA regulatory proteins. Methods Enzymol 429:299–321
4. Clement SL, Lykke-Andersen J (2008) A tethering approach to study proteins that activate mRNA turnover in human cells. Methods Mol Biol 419:121–133
5. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW (2004) Using the lambda N peptide to tether proteins to RNAs. Methods Mol Biol 257:135–154
6. Stubbs SH, Hunter OV, Hoover A, Conrad NK (2012) Viral factors reveal a role for REF/Aly in nuclear RNA stability. Mol Cell Biol 32(7):1260–1270
7. Graveley BR, Maniatis T (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell 1(5):765–771
8. Lim C et al (2011) The novel gene twenty-four defines a critical translational step in the Drosophila clock. Nature 470(7334):399–403
9. Lionnet T et al (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat Methods 8(2):165–170
10. Kim YK, Furic L, Desgroseillers L, Maquat LE (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3′ UTRs so as to elicit mRNA decay. Cell 120(2):195–208
11. Dugre-Brisson S et al (2005) Interaction of Staufen1 with the 5′ end of mRNA facilitates translation of these RNAs. Nucleic Acids Res 33(15):4797–4812
12. Ricci EP et al (2014) Staufen1 senses overall transcript secondary structure to regulate translation. Nat Struct Mol Biol 21(1):26–35
13. Kim J et al (2014) Splicing factor SRSF3 represses the translation of programmed cell death 4 mRNA by associating with the 5′-UTR region. Cell Death Differ 21(3):481–490
14. Stripecke R, Oliveira CC, McCarthy JE, Hentze MW (1994) Proteins binding to 5′ untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. Mol Cell Biol 14(9):5898–5909
15. De Gregorio E, Baron J, Preiss T, Hentze MW (2001) Tethered-function analysis reveals that elF4E can recruit ribosomes independent of its binding to the cap structure. RNA 7(1):106–113
16. Bardwell VJ, Wickens M (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic Acids Res 18(22):6587–6594
17. Pillai RS, Artus CG, Filipowicz W (2004) Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. RNA 10(10):1518–1525
18. Lykke-Andersen J, Shu MD, Steitz JA (2000) Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell 103(7):1121–1131
19. Gehring NH, Neu-Yilik G, Schell T, Hentze MW, Kulozik AE (2003) Y14 and hUpf3b form an NMD-activating complex. Mol Cell 11(4):939–949
20. Barreau C, Watrin T, Beverley Osborne H, Paillard L (2006) Protein expression is increased by a class III AU-rich element and tethered CUG-BP1. Biochem Biophys Res Commun 347(3):723–730
21. Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK (2005) The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. EMBO J 24(14):2656–2666
22. Gorgoni B et al (2005) The stem-loop binding protein stimulates histone translation at an early step in the initiation pathway. RNA 11(7):1030–1042
23. Williamson JR (2000) Induced fit in RNA-protein recognition. Nat Struct Biol 7(10):834–837
24. Frankel AD, Smith CA (1998) Induced folding in RNA-protein recognition: more than a simple molecular handshake. Cell 92(2):149–151

25. Grskovic M, Hentze MW, Gebauer F (2003) A co-repressor assembly nucleated by Sex-lethal in the 3′ UTR mediates translational control of Drosophila msl-2 mRNA. EMBO J 22(20):5571–5581
26. Sanford JR et al (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res 19(3):381–394
27. Ule J et al (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302(5648):1212–1215
28. Johansson HE et al (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. Proc Natl Acad Sci U S A 95(16):9244–9249
29. Tan R, Frankel AD (1995) Structural variety of arginine-rich RNA-binding peptides. Proc Natl Acad Sci U S A 92(12):5282–5286
30. Lim F, Peabody DS (2002) RNA recognition site of PP7 coat protein. Nucleic Acids Res 30(19):4138–4144
31. Weeks KM, Ampe C, Schultz SC, Steitz TA, Crothers DM (1990) Fragments of the HIV-1 Tat protein specifically bind TAR RNA. Science 249(4974):1281–1285
32. Goforth JB, Anderson SA, Nizzi CP, Eisenstein RS (2010) Multiple determinants within iron-responsive elements dictate iron regulatory protein binding and regulatory hierarchy. RNA 16(1):154–169
33. Katsamba PS, Myszka DG, Laird-Offringa IA (2001) Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. J Biol Chem 276(24):21476–21481
34. Witherell GW, Uhlenbeck OC (1989) Specific RNA-binding by Q beta coat protein. Biochemistry 28(1):71–76
35. Gott JM, Wilhelm LJ, Uhlenbeck OC (1991) RNA-binding properties of the coat protein from bacteriophage GA. Nucleic Acids Res 19(23):6499–6503
36. Carey J, Lowary PT, Uhlenbeck OC (1983) Interaction of R17 coat protein with synthetic variants of its ribonucleic acid binding site. Biochemistry 22(20):4723–4730
37. Lowary PT, Uhlenbeck OC (1987) An RNA mutation that increases the affinity of an RNA-protein interaction. Nucleic Acids Res 15(24):10483–10493
38. Romaniuk PJ, Lowary P, Wu HN, Stormo G, Uhlenbeck OC (1987) RNA-binding site of R17 coat protein. Biochemistry 26(6):1563–1568
39. Wu HN, Kastelic KA, Uhlenbeck OC (1988) A comparison of two phage coat protein-RNA interactions. Nucleic Acids Res 16(11):5055–5066
40. Carey J, Cameron V, de Haseth PL, Uhlenbeck OC (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. Biochemistry 22(11):2601–2610
41. Carey J, Uhlenbeck OC (1983) Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. Biochemistry 22(11):2610–2615
42. Lim F, Peabody DS (1994) Mutations that increase the affinity of a translational repressor for RNA. Nucleic Acids Res 22(18):3748–3752
43. Witherell GW, Wu HN, Uhlenbeck OC (1990) Cooperative binding of R17 coat protein to RNA. Biochemistry 29(50):11051–11057
44. Beck M et al (2011) The quantitative proteome of a human cell line. Mol Syst Biol 7:549
45. Peabody DS, Ely KR (1992) Control of translational repression by protein-protein interactions. Nucleic Acids Res 20(7):1649–1655
46. LeCuyer KA, Behlen LS, Uhlenbeck OC (1995) Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. Biochemistry 34(33):10600–10606
47. Peabody DS, Al-Bitar L (2001) Isolation of viral coat protein mutants with altered assembly and aggregation properties. Nucleic Acids Res 29(22), E113
48. Lazinski D, Grzadzielska E, Das A (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell 59(1):207–218
49. De Gregorio E, Preiss T, Hentze MW (1999) Translation driven by an eIF4G core domain in vivo. EMBO J 18(17):4865–4874
50. Austin RJ, Xia T, Ren J, Takahashi TT, Roberts RW (2002) Designed arginine-rich RNA-binding peptides with picomolar affinity. J Am Chem Soc 124(37):10966–10967

51. Van Gilst MR, Rees WA, Das A, von Hippel PH (1997) Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. Biochemistry 36(6):1514–1524
52. Van Gilst M, Rees WA, von Hippel PH (1995) Structural and thermodynamic characteristics of the binding of the lambda N protein to a RNA hairpin. Nucleic Acids Symp Ser 33:145–147
53. Lim F, Downey TP, Peabody DS (2001) Translational repression and specific RNA-binding by the coat protein of the Pseudomonas phage PP7. J Biol Chem 276(25):22507–22513
54. Carroll JS, Munchel SE, Weis K (2011) The DExD/H box ATPase Dhh1 functions in translational repression, mRNA decay, and processing body dynamics. J Cell Biol 194(4):527–537
55. 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(2):119–121
56. Wu B, Chen J, Singer RH (2014) Background free imaging of single mRNAs in live cells using split fluorescent proteins. Sci Rep 4:3615
57. Ponka P (1997) Tissue-specific regulation of iron metabolism and heme synthesis: distinct control mechanisms in erythroid cells. Blood 89(1):1–25
58. Thomson AM, Rogers JT, Leedman PJ (1999) Iron-regulatory proteins, iron-responsive elements and ferritin mRNA translation. Int J Biochem Cell Biol 31(10):1139–1152
59. Hentze MW, Kuhn LC (1996) Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. Proc Natl Acad Sci U S A 93(16):8175–8182
60. Pantopoulos K (2004) Iron metabolism and the IRE/IRP regulatory system: an update. Ann N Y Acad Sci 1012:1–13
61. Valegard K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature 371(6498):623–626
62. Rumnieks J, Tars K (2014) Crystal structure of the bacteriophage Qbeta coat protein in complex with the RNA operator of the replicase gene. J Mol Biol 426(5):1039–1049
63. Tars K, Bundule M, Fridborg K, Liljas L (1997) The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of Escherichia coli leviviruses. J Mol Biol 271(5):759–773
64. Lim F, Spingola M, Peabody DS (1996) The RNA-binding site of bacteriophage Qbeta coat protein. J Biol Chem 271(50):31839–31845
65. Ni CZ et al (1996) Crystal structure of the coat protein from the GA bacteriophage: model of the unassembled dimer. Protein Sci 5(12):2485–2493
66. Chen L, Frankel AD (1994) An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure. Biochemistry 33(9):2708–2715
67. Wakiyama M, Kaitsu Y, Muramatsu R, Takimoto K, Yokoyama S (2012) Tethering of proteins to RNAs using the bovine immunodeficiency virus-Tat peptide and BIV-TAR RNA. Anal Biochem 427(2):130–132
68. Tsai DE, Harper DS, Keene JD (1991) U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. Nucleic Acids Res 19(18):4931–4936
69. Tang J, Rosbash M (1996) Characterization of yeast U1 snRNP A protein: identification of the N-terminal RNA-binding domain (RBD) binding site and evidence that the C-terminal RBD functions in splicing. RNA 2(10):1058–1070
70. Brodsky AS, Silver PA (2000) Pre-mRNA processing factors are required for nuclear export. RNA 6(12):1737–1749
71. Takizawa PA, Vale RD (2000) The myosin motor, Myo4p, binds Ash1 mRNA via the adapter protein, She3p. Proc Natl Acad Sci U S A 97(10):5273–5278
72. Finoux AL, Seraphin B (2006) In vivo targeting of the yeast Pop2 deadenylase subunit to reporter transcripts induces their rapid degradation and generates new decay intermediates. J Biol Chem 281(36):25940–25947

73. van Gelder CW et al (1993) A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. EMBO J 12(13):5191–5200
74. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136(4):731–745
75. Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19(24):6860–6869
76. Le Hir H, Moore MJ, Maquat LE (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. Genes Dev 14(9):1098–1108
77. Singh G, Lykke-Andersen J (2003) New insights into the formation of active nonsense-mediated decay complexes. Trends Biochem Sci 28(9):464–466
78. Leeds P, Wood JM, Lee BS, Culbertson MR (1992) Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Mol Cell Biol 12(5):2165–2177
79. Lykke-Andersen J, Shu MD, Steitz JA (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science 293(5536):1836–1839
80. Fatscher T, Boehm V, Gehring NH (2015) Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. Cell Mol Life Sci 72(23):4523–4544
81. Wang X et al (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505(7481):117–120
82. Carmell MA, Xuan Z, Zhang MQ, Hannon GJ (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev 16(21):2733–2742
83. Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431(7006):343–349
84. Behm-Ansmant I et al (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev 20(14):1885–1898
85. Rehwinkel J, Behm-Ansmant I, Gatfield D, & Izaurralde E (2005) A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA (New York, N.Y.)* 11(11):1640–1647.
86. Lim C, Allada R (2013) ATAXIN-2 activates PERIOD translation to sustain circadian rhythms in Drosophila. Science 340(6134):875–879
87. Tacke R, Manley JL (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA-binding specificities. EMBO J 14(14):3540–3551
88. Shi H, Hoffman BE, Lis JT (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. Mol Cell Biol 17(5):2649–2657
89. Fu XD (1995) The superfamily of arginine/serine-rich splicing factors. RNA 1(7):663–680
90. Manley JL, Tacke R (1996) SR proteins and splicing control. Genes Dev 10(13):1569–1579
91. Caceres JF, Krainer AR (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. EMBO J 12(12):4715–4726
92. Zuo P, Manley JL (1993) Functional domains of the human splicing factor ASF/SF2. EMBO J 12(12):4727–4737
93. Wang J, Takagaki Y, Manley JL (1996) Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. Genes Dev 10(20):2588–2599
94. Tacke R, Chen Y, Manley JL (1997) Sequence-specific RNA-binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc Natl Acad Sci U S A 94(4):1148–1153
95. Chandler DS, Qi J, Mattox W (2003) Direct repression of splicing by transformer-2. Mol Cell Biol 23(15):5174–5185
96. Shen M, Mattox W (2012) Activation and repression functions of an SR splicing regulator depend on exonic versus intronic-binding position. Nucleic Acids Res 40(1):428–437

97. Sun S, Zhang Z, Fregoso O, Krainer AR (2012) Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. RNA 18(2):274–283
98. Zhang C et al (2008) Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. Genes Dev 22(18):2550–2563
99. Yeo GW et al (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol 16(2):130–137
100. Jansen RP, Dowzer C, Michaelis C, Galova M, Nasmyth K (1996) Mother cell-specific HO expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. Cell 84(5):687–697
101. Bobola N, Jansen RP, Shin TH, Nasmyth K (1996) Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. Cell 84(5):699–709
102. Long RM, Gu W, Lorimer E, Singer RH, Chartrand P (2000) She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. EMBO J 19(23):6592–6601
103. Bertrand E et al (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2(4):437–445
104. Mor A et al (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat Cell Biol 12(6):543–552
105. Nott A, Le Hir H, Moore MJ (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev 18(2):210–222
106. Buxbaum AR, Haimovich G, Singer RH (2015) In the right place at the right time: visualizing and understanding mRNA localization. Nat Rev Mol Cell Biol 16(2):95–109
107. Mao YS, Zhang B, Spector DL (2011) Biogenesis and function of nuclear bodies. Trends Genet 27(8):295–306
108. Engreitz J, Lander ES, Guttman M (2015) RNA antisense purification (RAP) for mapping RNA interactions with chromatin. Methods Mol Biol 1262:183–197
109. Engreitz JM et al (2014) RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 159(1):188–199
110. Gerstberger S, Hafner M, Ascano M, Tuschl T (2014) Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. Adv Exp Med Biol 825:1–55

Chapter 4
Single Molecule Approaches in RNA-Protein Interactions

Victor Serebrov and Melissa J. Moore

Abstract RNA-protein interactions govern every aspect of RNA metabolism, and aberrant RNA-binding proteins are the cause of hundreds of genetic diseases. Quantitative measurements of these interactions are necessary in order to understand mechanisms leading to diseases and to develop efficient therapies. Existing methods of RNA-protein interactome capture can afford a comprehensive snapshot of RNA-protein interaction networks but lack the ability to characterize the dynamics of these interactions. As all ensemble methods, their resolution is also limited by statistical averaging. Here we discuss recent advances in single molecule techniques that have the potential to tackle these challenges. We also provide a thorough overview of single molecule colocalization microscopy and the essential protein and RNA tagging and detection techniques.

Keywords CoSMoS • Single molecule imaging • TIRF • Protein tags • HILO

Abbreviations

CoSMoS	Colocalization single molecule spectroscopy
EM-CCD	Electron-multiplied charge-coupled device
HILO	Highly inclined and laminated optical sheet microscopy
RBP	RNA-binding protein
RNP	Ribonucleoprotein
sCMOS	Scientific complementary metal-oxide-semiconductor
TIRF	Total internal reflection fluorescence

V. Serebrov, Ph.D. • M.J. Moore, Ph.D. (✉)
Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA, USA
e-mail: melissa.moore@umassmed.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_4

1 Introduction

Ask any RNA scientist about his or her favorite molecule of life, and they (we included) could go on for hours about their/our fascination with RNA. RNA is a remarkably versatile biopolymer that, owing to a few key peculiarities of its chemical structure, can carry out nearly all the fundamental functions required for life: conveying genetic information, catalyzing chemical reactions and scaffolding macromolecular assemblages. Presumably, this multitude of functions made primitive life possible in the primordial RNA world. From the standpoint of highly evolved, protein-based modern organisms, however, there are remarkably few things that RNAs are now allowed to do on their own in their "naked" state. Instead, from the moment of their creation by an RNA polymerase until their demise at the hands of endo- and exonucleases, RNA molecules are almost continually "clothed" in proteins. Such RNA-binding proteins (RBPs) serve as key points of contact with the many cellular machineries vying for interaction with the underlying RNA [1, 2]. Some RBPs are tight-fitting and not easily shed, while others are more akin to readily exchangeable outerwear that can be tailored to suit the changing environment. Pre-mRNAs, for example, are co-transcriptionally clothed in proteins that direct their splicing and polyadenylation. While much of this clothing is shed along with the introns in the nucleus, the resultant mRNP is shipped to the cytoplasm outfitted with nuclear cap- and poly(A)-binding proteins and a tight-fitting undercoat of exon-junction complexes (EJCs) that help anchor in place a plethora of sequence-specific RBPs [3]. Once in the cytoplasm, mRNAs undergo dramatic wardrobe changes as the nuclear cap- and poly(A)-binding factors are replaced with their cytoplasmic counterparts, and the first round of translation strips off any proteins (including EJCs) associated with the opening reading frame. Translation efficiency is modulated by yet other RBPs bound to the UTRs. If a fault in translation is detected, due to a premature or absent stop codon or an obstruction in the ORF preventing ribosome read-through, then a specific decay pathway is activated via association with even more RBPs [4–6].

Even this exceedingly sketchy depiction of mRNP dynamics is sufficient to illustrate two of the main principles governing RNA-RBP interactions. The first principle is that an RNA's metabolism can be almost entirely described in terms of its associated RBPs. Since RNA function is largely mediated by its complement of bound proteins, RNP composition closely reflects the cellular process in which the RNA is engaged at any given time. For instance, the set of RBPs associated with an mRNA being actively translated will be consistent with polyribosomes, whereas those associated with a translationally silent mRNA will more likely characterize processing bodies or stress granules. Therefore, measuring and understanding the RNA-RBP interactome and how it changes in normal and diseased states is key if we are to fully understand the overlying cellular processes. Remarkably, recent mRNA interactome studies have identified >1500 different RBPs in human cells that interact directly with polyA+ RNA, a number reflecting the extreme complexity mRNA synthesis, function and decay. Mutations in a subset of these RBPs are linked to over

200 human diseases [7, 8], with neurological disorders, muscular atrophies, and cancer having the strongest documented ties to RBP malfunction [9, 10].

The second principle is that the RNA-protein interactome must be highly dynamic to keep up with the speed of cellular processes. In humans, the average mRNA half-life is a few hours, whereas in budding yeast it's only ~20 min. These numbers provide the floor for rates of mRNA interactome dynamics. The ceiling, however, is orders of magnitude faster. For instance, the entire time required for pre-mRNA maturation in the nucleus and export to the cytoplasm can be just a few minutes [11], and mRNAs subject to nonsense-mediated decay can have cytoplasmic half-lives of <1 min [12]. Within these short timeframes, many complicated processes must be initiated and completed, meaning that capturing the full dynamics of RNA-protein interactomes requires methods with time resolutions ranging from minutes down to fractions of seconds.

If we are to gain a systems level understanding of RNA-protein interactions, appropriate tools are needed to probe both RNP structure and RNP dynamics. Modern methods of interactome capture based on RNA-protein crosslinking followed by the mass spec identification of crosslinked proteins can provide a nearly complete snapshot of proteins bound to a starting pool of RNA molecules [7, 8]. Deep sequencing can also provide a transcriptome-wide picture of the exact sites of RNA-protein interaction [13]. Finally, recent development of robust methods for purifying endogenous RNPs now makes it possible to identify all the interaction partners of a single RNA of interest [14]. Together, these tools are rapidly expanding our knowledge of RNP parts lists and how the parts are distributed along RNA molecules. All such methods, however, rely on chemical or UV crosslinking to stabilize RNA-protein complexes. As crosslinking efficiencies can vary widely for different types of proteins, they may be a source of considerable bias that is exceedingly difficult to account for [13]. Further, structure-based methods cannot easily capture interaction dynamics. In addition to their relatively poor time resolution, the statistical averaging inherent to any ensemble technique makes it near impossible to decipher how different RBPs interact with one other within the context an intact RNP.

In this chapter, we discuss the advantages of direct single molecule imaging methodologies as applied to analysis of RBPs binding to RNA. We will overview basic principles, instrumentation and types of data that can be acquired. We will then cover our method of choice—Colocalization Single Molecule Microscopy (CoSMoS)—in more detail, as well as provide an in depth overview of the necessary protein and RNA labeling methodologies. As of this writing (late 2013) these methods have been applied to RNA-protein interactions either in completely recombinant, purified systems or in complex cell extracts [15–21]. But recent advances in single molecule imaging in live cells have already afforded the first glimpses at RNA-protein interactions at the single molecule level *in vivo* [22–26], and the same basic principles and considerations that we discuss here should be applicable to *in vivo* analyses as well. When combined with the structural data discussed above, the dynamics data provided by single molecule analysis hold the promise of providing a holistic picture of RNP metabolism.

2 Single Molecule Investigation of RNA-Protein Interactions

Although the field of single molecule biology is less than two decades old, single molecule techniques have already had a phenomenal impact on our understanding of biological systems and have firmly established themselves as an irreplaceable means of research. Here, we focus on their advantages in comparison with traditional "bulk" or "ensemble" approaches.

2.1 Why Look at Single Molecules?

A fundamental limitation of any technique that lumps together many molecules is that all measurements are subject to ensemble averaging, in which information about individual molecules is lost and replaced with a statistical average from which a mechanism can only be inferred, not directly observed. Due to such averaging, the field of chemical kinetics is a purely deductive one in which reaction schemes cannot be proven but can only be disproved. That is, different reaction mechanisms that produce identical kinetics cannot be distinguished. Further, many steps and branches are often undetectable. For consecutive reaction steps, only the slowest, rate-limiting step(s) will control the overall reaction rate. In a reaction with branched pathways where one branch is significantly faster than all others, the fastest process will largely determine the overall reaction kinetics and there may be no clear evidence of any slower, lesser-used pathways.

The problem of ensemble averaging is particularly challenging for the highly complex molecular mechanisms underlying most biological systems. For instance, molecular motors, such as polymerases, helicases, kinesins, myosins, dynein, the ribosome and other translocating enzymes, follow intricate kinetic mechanisms that may involve stochastic and systematic pausing, backtracking, slipping, directionality and rate switching, and other complex behaviors [27–29]. Single molecule techniques have revolutionized the field of molecular motors by enabling direct observation of these mechanistic features, many of which are completely hidden in traditional ensemble methods. Other important aspects of biological macromolecules are their dynamic and static heterogeneity [30]. Dynamic heterogeneity is a type of molecular memory leading to spontaneous fluctuations of activity by individual molecules [31]. Even modest enzymes with generally simple reaction mechanisms such as cholesterol oxidase have been shown to exhibit various degrees of dynamic heterogeneity [32]. Static heterogeneity implies that in each population of macromolecular species a fraction of molecules are completely inactive or have altered properties. Static heterogeneity may reflect either inherent heterogeneity such as incomplete post-translational protein modification or incomplete DNA methylation, or result from a multitude of spontaneous decay processes such as chemical modification (e.g., redox conversion of cysteine to cystine), chemical or enzymatic chain scission, or conformational changes (e.g., misfolding).

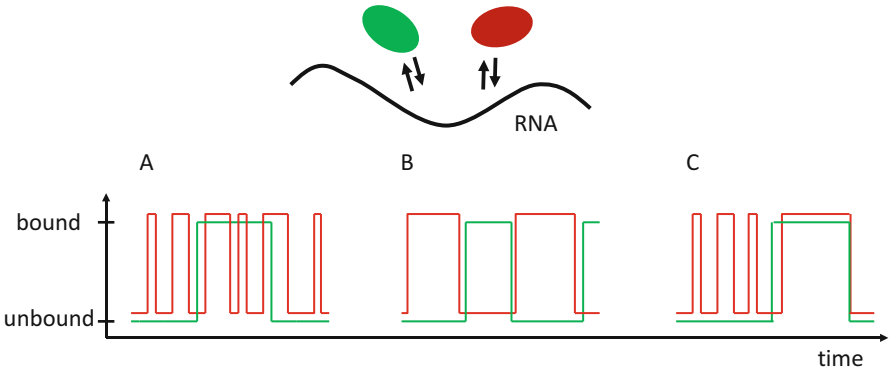


Fig. 4.1 Examples of single molecule trajectories showing various binding modes for two RBPs binding to the same RNA target. (**a**) uncorrelated binding; (**b**) mutually exclusive binding; (**c**) cooperative binding. The three examples illustrate different scenarios where binding of one protein is affected by binding of another protein. In the example (**a**), neither protein's binding affinity is affected by the presence of the other. These proteins therefore exhibit no cooperativity. Example (**b**) shows the extreme case of mutual exclusivity (negative cooperativity), where binding of one protein is prohibited if the other protein is already present. In example (**c**), binding affinity of one protein is increased when the other is already bound. In the most extreme case of such positive cooperativity, binding of one component can only occur after arrival of the other component; this results in highly ordered binding

Because single molecule methods monitor individual molecules and measure their properties directly, these altered or damaged populations can be easily discovered, and either accounted for or simply excluded from the analysis.

Some of the most fundamental questions regarding RNA-protein interactions center around the comings and goings of different proteins from a single RNA species. Figure 4.1 shows three simulations of single molecule trajectories for equilibrium binding of two RBPs to the same RNA molecule. In all three cases, the green and red proteins are bound roughly 1/3 and 2/3 of the time, respectively. Traditional ensemble assays can measure the equilibrium binding of both proteins and determine their relative affinities (or dissociation constants; K_d). In all cases shown, such analyses would correctly conclude that the red protein binds about twice as tightly as the green one. Due to statistical averaging, however, these assays will fail to reveal the different modes of protein binding, and therefore miss key aspects of the underlying mechanisms. Some of these aspects, such as mutual exclusivity or positive cooperativity, can be deduced from additional ensemble experiments, but to do this properly one needs to perform numerous titrations that are not always possible (e.g., when the sole source of both proteins is a single whole cell extract). If instead a single molecule technique that can detect individual interaction events were used, it would immediately be clear from just one experiment that the binding of the two proteins is completely uncorrelated in case A, that their binding is competitive and mutually exclusive in case B, and that binding of one protein is necessary to stabilize binding of the other in case C. Although only two colors are used in this illustration (which is an example of a two-color CoSMoS experiment, see below), four reaction

species exist: “naked” RNA, RNA with red protein bound, RNA with green protein bound, and RNA with both proteins bound. These distinct modes of protein binding can only be resolved if (1) all of these species are measured over time and (2) on the same RNA molecule. This is the sole domain of single molecule microscopy. Thus, this one simple example easily illustrates the many advantages of single molecule observation over traditional ensemble analyses.

2.2 Using Fluorescence Microscopy for Single Molecule Imaging

The most common way to visualize individual protein molecules is to make them fluorescent and observe them using a fluorescence microscope. Because a single fluorescent dye or protein is an extremely dim light source, a number of specialized approaches have been developed to gain a sufficiently high signal-to-background ratio for single molecule imaging. In general, this is achieved by increasing fluorophore brightness and/or reducing background fluorescence [33]. Although a fluorophore’s fluorescence yield is largely determined by its chemical structure, the number of photons (and therefore the signal) that can be extracted from a single dye is greatly improved by oxygen scavenging systems that eliminate the reactive oxygen species largely responsible for photodestruction processes [15, 34, 35]. Additionally, triplet state quenchers greatly minimize formation of long-lived non-fluorescent triplet states that result in dye blinking. Such quenchers therefore increase overall dye brightness [36, 37]. Finally, new brighter, more photostable and virtually pH-insensitive fluorescent dyes have been introduced in the last decade and are now commercially available, notably the AlexaFluor and DyLight families of dyes.

A number of illumination techniques have been developed that aim to limit illumination to a very small volume, and therefore reduce the out-of-focus fluorescence, which is a major drawback of traditional widefield illumination (Fig. 4.2a). In the widefield setup, the entire specimen thickness (or solution) is illuminated and produces fluorescence that is collected by the microscope objective. The bright out-of-focus fluorescence becomes a fluorescent background that makes detection of a single fluorophore all but impossible. One way to minimize this background fluorescence is to reject it using a narrow aperture (a pinhole) for both the excitation and emission light, a technique known as confocal microscopy (Fig. 4.2d). Scanning confocal microscopy is widely used for cell imaging. Because a pinhole is used in the emission path, however, the confocal setup collects only a small fraction of fluorescence emission resulting in a relatively low sensitivity not suitable for single fluorophore imaging. Notable exceptions are the non-imaging techniques such as fluorescent correlation and cross-correlation spectroscopy (FCS and FCCS). In FCS and its variants, the emission fluorescence from the entire confocal volume (<1 femtoliter, or 10^{-15} L) is measured by an avalanche photodiode, followed by analyses of temporal autocorrelation or cross-correlation.

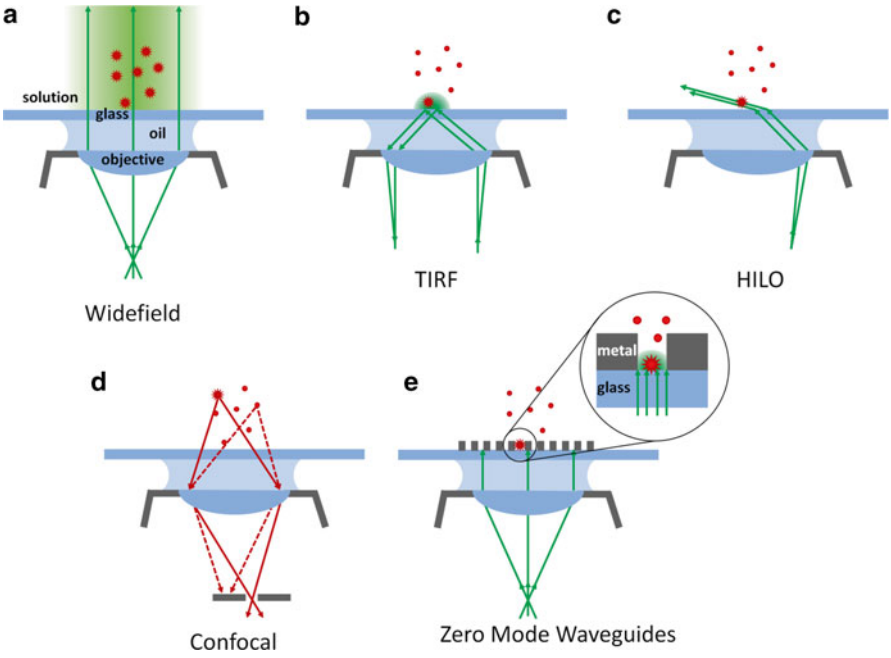


Fig. 4.2 Illumination and imaging schemes used in cell and single molecule imaging. (**a**) Conventional widefield configuration illuminates the entire specimen thickness, making it unsuitable for single molecule imaging due to high levels of fluorescent background. TIRF (**b**) and HILO (**c**) are two illumination techniques that aim to restrict the illuminated volume in order to reduce out-of-focus background fluorescence. The depth of illumination is a few micrometers in HILO and about a hundred nanometers in TIRF. The reduced fluorescent background makes both techniques suitable for single molecule imaging. The advantage of HILO is its ability to image inside cells. TIRF has superior signal-to-noise characteristics but is restricted to imaging directly on or very close to the slide surface. (**d**) Confocal microscopy is a scanning imaging technique utilizing a pinhole in the path of both excitation and emission light (emission is shown) in order to reject out-of-focus background. (**e**) Zero mode waveguides (ZMWs) feature zeptoliter-volume wells in a thin layer of aluminum atop a glass slide. The diameter of each well (~50 nm) is smaller than the excitation light wavelength. Due to the metal walls, excitation light in ZMWs can only propagate a few tens of nanometers inside the well

Building on advances in nanofabrication, one of the latest developments takes advantage of the behavior of light when it is physically confined to a space smaller than its wavelength. Zero-mode waveguides (ZMWs) are nanofabricated wells holding mere zeptoliters (10^{-21} L) in a ~100 nm thick aluminum film deposited on a microscope slide (Fig. 4.2e). The well diameter is ~50 nm, smaller than the wavelength of visible light; excitation light cannot propagate in a channel this narrow so instead quickly decays within a few tens of nanometers of the glass surface [38]. Because of the extremely small illuminated volume, the background fluorescence is low enough to image a single fluorescent dye inside a ZMW well even when the solution above the well contains ~50 μM fluorescent molecules. For comparison, in

TIRF imaging the concentration of a fluorescent component in solution cannot exceed ~100 nM [39, 40]. ZMWs have already proven an invaluable imaging technique for applications where high concentrations (tens of μM or higher) of fluorescent molecules are required, and therefore reduction of fluorescent background becomes critical [41–43].

Due to its superior signal-to-noise ratio and relatively simple instrumentation, most *in vitro* (i.e., in cell extracts or recombinant purified systems) single molecule fluorescence studies to date have utilized total internal reflection fluorescence (TIRF, Fig. 4.2b) [28, 44, 45]. TIRF utilizes an excitation beam that is directed at an angle to the slide surface exceeding the critical angle at which light undergoes total internal reflection from the interface between two media with different refraction indices (total internal reflection only occurs if light travels from a medium with a higher refraction index into a medium with lower refraction index). Because of the wave properties of light, it cannot simply "stop" at the interface. Instead, the light "bleeds through" creating an evanescent illumination field on the other side of the interface, i.e., the slide surface. The evanescent field only illuminates a layer about 100 nm thick immediately adjacent to the slide surface. TIRF is therefore limited to surface imaging, with live cell imaging using TIR illumination being restricted to cell surface. A related illumination technique termed HILO (Fig. 4.2c) for highly inclined and laminated optical sheet has been used for single fluorophore imaging inside living cells [46–48]. HILO takes advantage of an optical setup similar of that of the objective-type TIRF, where a steep inclination angle of the excitation beam is achieved through use of a high numerical aperture objective.

3 Single Molecule Co-localization Spectroscopy (CoSMoS)

Together with Jeff Gelles' group (Brandeis University), we have spent the last several years developing methodologies to study the assembly and dynamics of macromolecular complexes using multi-color, single molecule co-localization. Termed CoSMoS, for Co-localization of Single Molecules Spectroscopy, this approach is based on direct observation of interactions between macromolecules using single molecule TIRF (SM-TIRF) [19]. In CoSMoS, interacting components of a macromolecular complex are labeled with differently colored fluorescent dyes, with one of the components being immobilized on the slide surface of a multi-wavelength TIRF microscope. The interactions between macromolecules are detected in real time as co-localization of different colored fluorescent spots. The number of components that can be imaged simultaneously is limited by available combinations of fluorescent dyes with essentially non-overlapping excitation and emission spectra as well as available laser lines, which together currently restrict most experiments to three colors.

CoSMoS has emerged as a powerful and unique single molecule tool to quantitatively measure complex interactions between macromolecules. It has been equally useful when used in purified systems reconstituted from recombinant components [20, 21], and in non-purified systems such as whole cell extracts [15–17]. CoSMoS

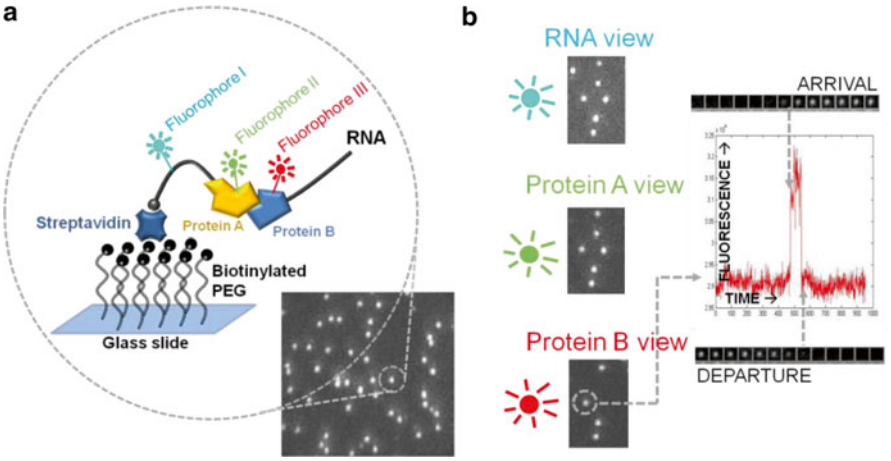


Fig. 4.3 (**a**) In this example of a typical CoSMoS experiment, interactions of a fluorescent, surface immobilized RNA with two fluorescent proteins are continuously monitored using three laser wavelengths. (**b**) Co-localization of fluorescent signals from individual RNA and protein molecules indicates binding events (simulated data from [19] are shown). Fluorescent emission signals from all interacting partners are recorded separately and simultaneously. Complete kinetic information, including on and off rates, can be extracted from single molecule fluorescence traces. Because individual molecules are monitored in real time, transient interactions and rare/alternative kinetic pathways can be detected

has the capacity to provide a complete kinetic description of interactions between multiple species of macromolecules and therefore elucidate the underlying molecular mechanisms in explicit detail. Figure 4.3 illustrates a typical CoSMoS experiment where binding of two proteins to the same RNA molecule is monitored over time. The proteins need to be tagged with mutually orthogonal tags that allow them to be labeled with different color fluorescent dyes in the same cell extract. For recombinant proteins, conventional label conjugation techniques can be used [49]. Fluorescent proteins of different colors can also be used but make imaging more difficult due to their relatively poor photophysical properties (see below). In this example, the RNA and two proteins are labeled with three spectrally distinguishable fluorescent dyes, and RNA is biotinylated and attached to the microscope slide surface via interaction with surface-bound streptavidin. Individual molecules of each species are detected as diffraction-limited spots of fluorescence. Images in each of the three colors are separated by the microscope optics and recorded over the entire time of the experiment. The slide surface is coated with a dense brush of high molecular weight polyethylene glycol (PEG) to prevent non-specific protein sticking [50]. The locations of RNA molecules on the surface are determined from the acquired RNA images, and then mapped onto the protein A and protein B images. As shown for protein B in Fig. 4.3, single molecule trajectories of protein binding and dissociation events are then recorded by quantification of protein fluorescence at the RNA locations.

An example of a real experiment following the association of U1, U2 and U5 snRNPs coming and going from individual pre-mRNA molecules in whole cell extract is shown in Fig. 4.4. This experiment unveils important mechanistic features

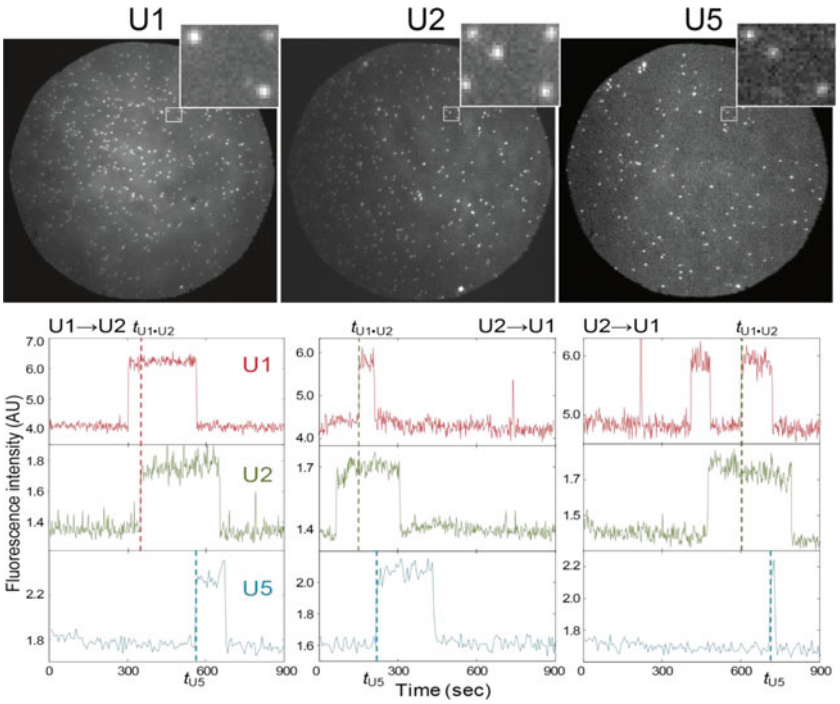


Fig. 4.4 An example of a CoSMoS experiment monitoring spliceosome assembly on individual pre-mRNAs from [17]. Top, three fields of view corresponding to fluorescent emission of U1, U2 and U5 snRNPs. Bottom, examples of single molecule trajectories for binding of U1, U2 and U5 snRNPs to three individual pre-mRNA molecules

of spliceosome assembly [17]. U1 and U2 snRNP can bind to the pre-mRNA in any order (U1 first or U2 first), reflecting their binding to separate sites on the pre-mRNA. The resulting complex is functionally identical regardless of the order of arrival. As clearly seen from the traces shown, U5 binding (as well as all following steps) requires that both U1 and U2 be already bound. In addition, this experiment distinctly demonstrates that all binding steps are reversible. As discussed in detail above, such complete and detailed assessment of a macromolecular assembly would have been unattainable in "bulk" experiments.

3.1 CoSMoS Instrumentation

The essential part of CoSMoS is a TIRF microscope equipped with lasers compatible with the excitation spectra of the fluorescent dyes being used. The imaging path of the microscope system should have dichroic splitters and emission filters that can separate fluorescence emitted by different color fluorescent dyes [50]. Because single fluorophores are extremely dim objects, it is typically necessary to use either an

electron-multiplied charge-coupled device (EM-CCD) camera or a scientific complementary metal-oxide-semiconductor (sCMOS) camera to record images, which are two of the best available choices for low light sensitivity. A lab-built CoSMoS microscope has been described by the Gelles group and this design remains the best option for multi-color CoSMoS [19]. However, all of the major microscope manufacturers (Leica, Nikon, Olympus and Zeiss) now offer single-molecule capable TIRF microscope systems with multi-color laser illumination. With proper selection of excitation and emission filters, these systems can perform well for many CoSMoS applications. Further, existing inverted microscopes can often be retrofitted for single molecule TIRF/CoSMoS applications by adding a TIRF illuminator, appropriate lasers and an EM-CCD or sCMOS camera. The cost of a commercial CoSMoS-capable system, while substantial, is within the means of many microscopy core facilities and individual laboratories.

3.2 CoSMoS: Practical Considerations

The main advantage of direct single molecule imaging is the ability to resolve what two or more individual binding partners are doing with respect to each other, and thus determine how they function within the context of a larger macromolecular assembly. In the case of CoSMoS and similar colocalization techniques, this means that it is necessary to ensure quantitative or near-quantitative fluorescent labeling for all interacting partners whose binding is to be detected in the colocalization experiment. If a significant fraction of unlabelled, non-fluorescent protein is present, this can lead to ambiguities in interpretation regarding its binding mechanism, as is illustrated with the examples in Fig. 4.1. For this reason, the choice of labeling techniques is critical. We have had the greatest success to date with the highly specific self-labeling protein tags (SNAP, CLIP, and ecDHFR, see below). These tags can be introduced genetically, and are one of few labeling strategies available when the biological process of interest needs to be carried out in cell extracts, since specific labeling needs to be performed in the background of thousands other proteins. They also offer a superior alternative when compared to traditional chemical conjugation methods used to fluorescently label recombinant proteins, which can often lead to loss of activity or functional heterogeneity of the labeled protein.

Another important consideration (especially for TIRF) is the normal functioning concentration of the protein of interest. As shown on Fig. 4.4, protein abundance in budding yeast spans about four orders of magnitude [51]. Some components of the translation machinery, which constitutes much of the protein content of living cells, can exceed several hundred thousand molecules per cell. This level of expression translates into intracellular concentrations >10 μM. On the other hand, DNA replication factors are often present at the levels of less than a hundred molecules per cell, which translates to nanomolar concentrations. Since the concentration of the protein is intimately linked to its binding affinity in a macromolecular assembly, the highly expressed proteins often have K_d's in the μM range (i.e., weak binding affinity);

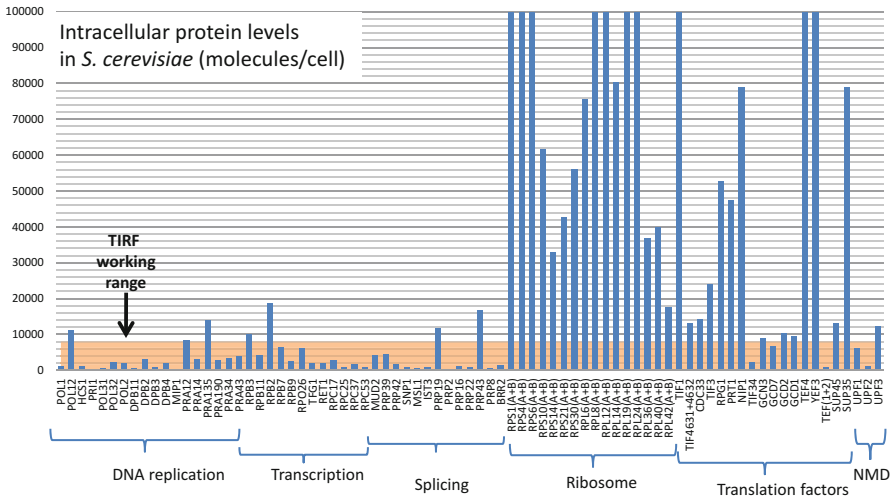


Fig. 4.5 The expression levels of proteins of interest are an important parameter to consider in CoSMoS experiments [51]. In CoSMoS, single molecule imaging utilizes a TIRF illumination setup in order to reduce background fluorescence. Even with TIRF illumination, however, the concentration of fluorescent protein must be kept less than 50–100 nM to achieve sufficiently high signal-to-noise ratio necessary to see single fluorescent dyes. In yeast extracts, this limits the abundance of fluorescently tagged proteins to less than about 10,000 molecules/cell

this in turn necessitates keeping them at high concentration to maintain their biological function, as is the case for many translation factors. The latter becomes a problem when single molecule fluorescence imaging is desired. Although TIR illumination greatly reduces background fluorescence, the practical concentration limit for a fluorescent protein cannot exceed 100 nM. We semi-empirically determined that in case of *S. cerevisiae* whole cell extracts, this limit translates to protein abundances of 8000–10,000 molecules per cell. While many proteins involved in DNA replication, transcription and splicing are below this limit, and therefore amenable to CoSMoS analysis, the endogenous levels of most ribosomal proteins and other translation factors would be problematic for CoSMoS experiments (Fig. 4.5).

3.3 Fluorescent Proteins

The main advantage of fluorescent proteins (FPs) as tags is their intrinsic fluorescence can be genetically encoded, therefore eliminating the need for the labeling step and circumventing its potential caveats, such as incomplete and non-specific labeling from which all other tags suffer. FPs are relatively small (~27 kDa) and stable barrel-like fold proteins with little or no affinity (i.e., stickiness) for other cellular proteins. The source their fluorescence is a fluorophore formed by oxidative cyclization of three amino acid side chains inside the barrel [52]. A palette of

fluorescent proteins is now available, with excitation wavelengths ranging from 400 to 600 nm [53]. Although the blue and green FPs are generally the brightest, this added signal is often negated by cellular autofluorescence (also present in cell extracts), which is most pronounced in the blue-green region. In comparison with organic fluorophores, however, fluorescent proteins have poor photochemical characteristics: they are relatively dim, prone to blinking and have relatively low photostability, making them less than ideal for single molecule imaging [54]. Nevertheless, the benefits of genetically encoded fluorescence often outweigh these drawbacks. Recently, novel fluorescent proteins were engineered based on bacterial phytochromes that are near-infrared emitters and appear to have better photochemical properties than GFP-derived FPs [55].

3.4 SNAP and CLIP Tags

SNAP and CLIP tags are small (~20 kDa) protein tags derived from the DNA-repair protein O^6-alkylguanine-DNA alkyltransferase [56]. This enzyme repairs O^6-alkylated guanine by transferring the alkyl group onto itself. When used as a fusion with a protein of interest, this protein tag will rapidly react with derivatives of benzyl-guanine under physiological conditions. This results in a covalent bond between the tag and the benzyl group, and therefore attachment of any small molecule linked to the benzyl group, such as a fluorescent dye. The CLIP tag is a modified version of SNAP that has been engineered to react with benzylcytosine while exhibiting virtually no reactivity towards benzylguanine. This essentially makes SNAP and CLIP orthogonal tags that can be used simultaneously, i.e. to label two proteins in the same cells or cell extract with different color fluorescent dyes [57]. Recently, the SNAP and CLIP tags have been further optimized to allow for more rapid labeling characteristics. Labeling rates of 10^4–10^5 $M^{-1}s^{-1}$ can be achieved with these new versions (termed $SNAP_f$ and $CLIP_f$), which translates into mere minutes to complete labeling at micromolar concentration of a benzylguanine (or benzylcytosine) derivative [58]. Nonetheless, it should be noted that $CLIP_f$ tag labeling is about an order of magnitude slower than $SNAP_f$ tag labeling.

3.5 HaloTag

The 33 kDa HaloTag (Promega) is based on an engineered version of a bacterial haloalkane dehalogenase enzyme. This enzyme removes halogens (e.g., chlorine) from aliphatic hydrocarbons. This reaction occurs via a two-step mechanism, wherein the first step results in an intermediate that has the aliphatic alkyl group covalently attached to an aspartate in the enzyme active site. During the second step this intermediate is hydrolyzed to recycle the enzyme via a water molecule activated by a catalytic histidine residue. The HaloTag is a mutant version in which this

histidine is replaced with phenylalanine and therefore is unable to carry out the second step. When used as a tag and fused to a protein of interest, the HaloTag allows for covalent attachment of any small molecule linked to the alkyl group of an alkylhalide. It has been reported that the labeling rate of the HaloTag is in the order of 10^6 $M^{-1}s^{-1}$, which is 10–100 times faster than that of $SNAP_f$ and $CLIP_f$ tags, and appears to be the fastest covalent labeling attainable with any protein tag [59, 60]. HaloTag is essentially bio-orthogonal in eukaryotes, showing a high degree of specificity in labeling reactions *in vivo* and *in vitro*.

3.6 DHFR Tag

The use of *E. coli* dihydrofolate reductase (ecDHFR) as a tag for non-covalent labeling was pioneered by the group of Virginia Cornish at Columbia University [61]. DHFR catalyzes NADPH-dependent conversion of dihydrofolate into tetrahydrofolate, a reaction important for the *de novo* synthesis of thymine. *E. coli* DHFR is a natural target of the antibiotic trimethoprim (TMP). TMP has fairly low affinity for mammalian DHFR, but it binds to ecDHFR with subnanomolar affinity in the presence of NADPH [62]. This interaction is stable enough to make the ecDHFR tag generally suitable for *in vivo* and *in vitro* imaging, including imaging of single molecules. The imaging can be carried out in the presence of a low concentration of TMP conjugated to a fluorescent dye (e.g., Cy3-TMP), which will bind tightly to the ecDHFR tag and render it fluorescent. We have successfully employed this strategy for labeling snRNPs in single molecule studies of spliceosome assembly. The advantages of the DHFR tag are twofold. First, it is suitable for imaging schemes that require very long acquisition times by largely circumventing the fluorophore photobleaching problem—because of the reversible nature of TMP binding, a photobleached fluorophore will eventually dissociate and be replaced with a new one. Second, as with FPs, there is no need for a covalent labeling step that inevitably requires post-labeling removal of free dye. However, the latter is also a disadvantage of the ecDHFR tag—since free dye-TMP needs to be present at all times during imaging, it contributes to background fluorescence. For TIRF imaging, this limits its concentration to well below 100 nM; this in turn limits the concentration of the protein of interest if its quantitative labeling is desired. Another concern with the edDHFR tag is that it is not guaranteed to have a dye-TMP molecule bound at all times (although the latter problem can be largely avoided by having more than one ecDHFR tag on the same protein or macromolecular complex).

4 CoSMoS Data Analysis

As shown by the examples in Figs. 4.1 and 4.3, the principal readout of a CoSMoS experiment are single molecule binding trajectories that are records of arrivals and departures of the labeled species on a surface-tethered molecule. Therefore, the data

that can be acquired for individual molecules are the arrival times (the time it takes for the protein to bind; e.g. to an RNA molecule from the moment of reagent mixing or from the last departure event), and the dwell time (the binding time spanned between arrival and departure). The statistics for both of these event types are described by a binomial distribution (in simple terms, only two outcomes or "states" are possible, bound or not). The standard deviation in such experiments rapidly decreases with increasing number of observations ($SD \sim 1/\sqrt{N}$) [63]. Therefore, it is typically sufficient to measure arrival and departure times on just a few hundreds of molecules (or binding events) to obtain satisfactory statistics. In contrast to the number of events, the rates of events (i.e., arrival and dwell times) follow exponential distributions [28, 30]. The principal feature of an exponential distribution is that the probability of an event (either arrival or departure in our case) is proportional to the waiting time, where the proportionality coefficient is the kinetic rate constant. An exponential distribution is a single parameter distribution, and therefore is explicitly defined by the rate constant. Rate constants for binding and dissociation reactions can be determined from the distribution parameters of arrival and dwell times, respectively [64]. When both on- (k_{on}) and off-rate (k_{off}) constants are known, this immediately yields the dissociation constant K_d, since $K_d = k_{off}/k_{on}$. As discussed before, different binding scenarios (such as those shown in Fig. 4.1) can be discovered by this analysis and quantitatively described.

It is important to note, however, that because CoSMoS utilizes fluorescence, an apparent departure event may in fact be a photobleaching or photodestruction event of the fluorescent dye. Photobleaching does not affect detection of arrival events, and thus on-rate measurements are much easier to determine. To accurately determine off-rates, photobleaching must be either minimized below a negligible level, or the photobleaching rate must be measured and accounted for. In the latter case the experiments need to be repeated at multiple laser power levels and the actual departure rate found by extrapolating to zero laser power.

5 Conclusions

Single molecule imaging techniques have a clear advantage over traditional ensemble methods in that they provide direct and explicit information as to macromolecular assembly and dynamics. Although their use for the analysis of RNA-protein interactions and other interactomes is in its infancy, the potential to bring the power of single molecule inquiry to complex interaction networks has already been well demonstrated. In the last decade we have witnessed considerable progress in both single molecule microscopy and specific protein and nucleic acid labeling methodologies, which has established a strong foundation for future progress. One can expect that further developments in these fields, and particularly their application *in vivo* and/or on a transcriptome-wide scale, will generate new great tools capable of dissecting complex macromolecular interactions in unprecedented detail on an unprecedented scale.

References

1. Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science 309:1514–1518. doi:10.1126/science.1111443
2. Balagopal V, Parker R (2009) Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. Curr Opin Cell Biol 21:403–408. doi:10.1016/j.ceb.2009.03.005
3. Singh G, Kucukural A, Cenik C et al (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell 151:750–764. doi:10.1016/j.cell.2012.10.007
4. Graille M, Séraphin B (2012) Surveillance pathways rescuing eukaryotic ribosomes lost in translation. Nat Rev Mol Cell Biol 13:727–735. doi:10.1038/nrm3457
5. Nürenberg E, Tampé R (2013) Tying up loose ends: ribosome recycling in eukaryotes and archaea. Trends Biochem Sci 38:64–74. doi:10.1016/j.tibs.2012.11.003
6. Parker R (2012) RNA degradation in Saccharomyces cerevisae. Genetics 191:671–702. doi:10.1534/genetics.111.137265
7. 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. doi:10.1016/j.molcel.2012.05.021
8. 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. doi:10.1016/j.cell.2012.04.031
9. Castello A, Fischer B, Hentze MW, Preiss T (2013) RNA-binding proteins in Mendelian disease. Trends Genet 29:318–327. doi:10.1016/j.tig.2013.01.004
10. Lukong KE, Chang K, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends Genet 24:416–425. doi:10.1016/j.tig.2008.05.004
11. Audibert A, Weil D, Dautry F (2002) In Vivo kinetics of mRNA splicing and transport in mammalian cells. Mol Cell Biol 22(19):6706–6718. doi:10.1128/MCB.22.19.6706
12. Trcek T, Sato H, Singer RH, Maquat LE (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. Genes Dev 27:541–551. doi:10.1101/gad.209635.112
13. Singh G, Ricci EP, Moore MJ (2013) RIPiT-Seq: a high-throughput approach for footprinting RNA: protein complexes. Methods 65(3):320–332. doi:10.1016/j.ymeth.2013.09.013
14. Chu C, Qu K, Zhong FL et al (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44:667–678. doi:10.1016/j.molcel.2011.08.027
15. Crawford DJ, Hoskins AA, Friedman LJ et al (2008) Visualizing the splicing of single pre-mRNA molecules in whole cell extract. RNA 14:170–179. doi:10.1261/rna.794808
16. Hoskins AA, Friedman LJ, Gallagher SS et al (2011) Ordered and dynamic assembly of single spliceosomes. Science 331:1289–1295. doi:10.1126/science.1198830
17. Shcherbakova I, Hoskins AA, Friedman LJ et al (2013) Alternative spliceosome assembly pathways revealed by single-molecule fluorescence microscopy. Cell Rep 5:151–165. doi:10.1016/j.celrep.2013.08.026
18. Hoskins AA, Gelles J, Moore MJ (2011) New insights into the spliceosome by single molecule fluorescence microscopy. Curr Opin Chem Biol 15:864–870. doi:10.1016/j.cbpa.2011.10.010
19. Friedman LJ, Chung J, Gelles J (2006) Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence. Biophys J 91:1023–1031. doi:10.1529/biophysj.106.084004
20. Friedman LJ, Mumm JP, Gelles J (2013) RNA polymerase approaches its promoter without long-range sliding along DNA. Proc Natl Acad Sci U S A 110:9740–9745. doi:10.1073/pnas.1300221110
21. Friedman LJ, Gelles J (2012) Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. Cell 148:679–689. doi:10.1016/j.cell.2012.01.018
22. Grünwald D, Singer RH (2010) In Vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature 467:604–607. doi:10.1038/nature09438

23. Grünwald D, Singer RH, Rout M (2011) Nuclear export dynamics of RNA-protein complexes. Nature 475:333–341. doi:10.1038/nature10318
24. Manley S, Gillette JM, Patterson GH et al (2008) High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat Methods 5:155–157. doi:10.1038/NMETH.1176
25. Pinaud F, Dahan M (2011) Targeting and imaging single biomolecules in living cells by complementation-activated light microscopy with split-fluorescent proteins. Proc Natl Acad Sci U S A 108:E201–E210. doi:10.1073/pnas.1101929108
26. Xia T, Li N, Fang X (2013) Single-molecule fluorescence imaging in living cells. Annu Rev Phys Chem 64:459–480. doi:10.1146/annurev-physchem-040412-110127
27. Myong S, Ha T (2010) Stepwise translocation of nucleic acid motors. Curr Opin Struct Biol 20:121–127. doi:10.1016/j.sbi.2009.12.008
28. Tinoco I, Gonzalez RL (2011) Biological mechanisms, one molecule at a time. Genes Dev 25:1205–1231. doi:10.1101/gad.2050011
29. Bustamante C, Cheng W, Mejia YX, Meija YX (2011) Revisiting the central dogma one molecule at a time. Cell 144:480–497. doi:10.1016/j.cell.2011.01.033
30. Shi J, Dertouzos J, Gafni A, Steel D (2008) Application of single-molecule spectroscopy in studying enzyme kinetics and mechanism. Methods Enzymol 450:129–157. doi:10.1016/S0076-6879(08)03407-1
31. Min W, Jiang L, Xie XS (2010) Complex kinetics of fluctuating enzymes: phase diagram characterization of a minimal kinetic scheme. Chem Asian J 5:1129–1138. doi:10.1002/asia.200900627
32. Lu HP, Xun L, Xie XS (1998) Single-molecule enzymatic dynamics. Science 282:1877–1882
33. Walter NG, Huang C, Manzo AJ, Sobhy MA (2008) Do-it-yourself guide: how to use the modern single-molecule toolkit. Nat Methods 5:475–489. doi:10.1038/NMETH.1215
34. Aitken CE, Marshall RA, Puglisi JD (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys J 94:1826–1835. doi:10.1529/biophysj.107.117689
35. Vogelsang J, Kasper R, Steinhauer C et al (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. Angew Chem Int Ed Engl 47:5465–5469. doi:10.1002/anie.200801518
36. Rasnik I, Mckinney SA, Ha T (2006) Nonblinking and long- lasting single-molecule fluorescence imaging. Nat Methods 3:891–893. doi:10.1038/NMETH934
37. Dave R, Terry DS, Munro JB, Blanchard SC (2009) Mitigating unwanted photophysical processes for improved single-molecule fluorescence imaging. Biophys J 96:2371–2381. doi:10.1016/j.bpj.2008.11.061
38. Levene MJ, Korlach J, Turner SW et al (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. Science 299:682–686. doi:10.1126/science.1079700
39. Leslie SR, Fields AP, Cohen AE (2010) Convex lens-induced confinement for imaging single molecules. Anal Chem 82:6224–6229. doi:10.1021/ac101041s
40. Elting MW, Leslie SR, Churchman LS et al (2013) Single-molecule fluorescence imaging of processive myosin with enhanced background suppression using linear zero-mode waveguides (ZMWs) and convex lens induced confinement (CLIC). Opt Express 21:1189–1202. doi:10.1364/OE.21.001189
41. Uemura S, Aitken CE, Korlach J et al (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. Nature 464:1012–1017. doi:10.1038/nature08925
42. Petrov A, Kornberg G, O'Leary S et al (2011) Dynamics of the translational machinery. Curr Opin Struct Biol 21:137–145. doi:10.1016/j.sbi.2010.11.007
43. Chen J, Dalal RV, Petrov AN et al (2013) High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. Proc Natl Acad Sci U S A 111(2):664–669. doi:10.1073/pnas.1315735111/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1315735111

44. Axelrod D (2003) Total internal reflection fluorescence microscopy in cell biology. Methods Enzymol 361:1–33
45. Axelrod D (2013) Evanescent excitation and emission in fluorescence microscopy. Biophys J 104:1401–1409. doi:10.1016/j.bpj.2013.02.044
46. Tokunaga M, Imamoto N, Sakata-sogawa K (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5:159–161. doi:10.1038/NMETH.1171
47. Pitchiaya S, Androsavich JR, Walter NG (2012) Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. EMBO Rep 13:709–715. doi:10.1038/embor.2012.85
48. Pitchiaya S, Krishnan V, Custer TC, Walter NG (2013) Dissecting non-coding RNA mechanisms in cellulo by single-molecule high-resolution localization and counting. Methods 63:188–199. doi:10.1016/j.ymeth.2013.05.028
49. Hermanson GT (2008) Bioconjugation techniques. Acad Press 10:0123705010. doi: 10.1007/s00216-009-2731-y
50. Roy R, Hohng S, Ha T (2008) A practical guide to single-molecule FRET. Nat Methods 5:507–516. doi:10.1038/nmeth.1208
51. Ghaemmaghami S, Huh W-K, Bower K et al (2003) Global analysis of protein expression in yeast. Nature 425:737–741. doi:10.1038/nature02046
52. Remington SJ (2006) Fluorescent proteins: maturation, photochemistry and photophysics. Curr Opin Struct Biol 16:714–721. doi:10.1016/j.sbi.2006.10.001
53. Stepanenko OV, Stepanenko OV, Shcherbakova DM et al (2011) Modern fluorescent proteins: from chromophore formation to novel intracellular applications. Biotechniques 51:313–314. doi:10.2144/000113765, 316, 318 passim
54. Ha T, Tinnefeld P (2012) Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. Annu Rev Phys Chem 63:595–617. doi:10.1146/annurev-physchem-032210-103340
55. Shcherbakova DM, Verkhusha VV (2013) Near-infrared fluorescent proteins for multicolor in vivo imaging. Nat Methods 10:751–754. doi:10.1038/nmeth.2521
56. Juillerat A, Gronemeyer T, Keppler A et al (2003) Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. Chem Biol 10:313–317
57. Gautier A, Juillerat A, Heinis C et al (2008) An engineered protein tag for multiprotein labeling in living cells. Chem Biol 15:128–136. doi:10.1016/j.chembiol.2008.01.007
58. Sun X, Zhang A, Baker B et al (2011) Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. Chembiochem 12:2217–2226. doi:10.1002/cbic.201100173
59. Los GV, Encell LP, McDougall MG et al (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3:373–382. doi:10.1021/cb800025k
60. Los GV, Wood K (2007) The HaloTag: a novel technology for cell imaging and protein analysis. Methods Mol Biol 356:195–208
61. Miller LW, Cai Y, Sheetz MP, Cornish VW (2005) In Vivo protein labeling with trimethoprim conjugates a flexible chemical tag. Nat Methods 2:255–257. doi:10.1038/NMETH749
62. Calloway NT, Choob M, Sanz A et al (2007) Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. Chembiochem 8:767–774. doi:10.1002/cbic.200600414
63. Floyd DL, Harrison SC, van Oijen AM (2010) Analysis of kinetic intermediates in single-particle dwell-time distributions. Biophys J 99:360–366. doi:10.1016/j.bpj.2010.04.049
64. Moffitt JR, Chemla YR, Bustamante C (2010) Methods in statistical kinetics. Methods Enzymol 475:221–257. doi:10.1016/S0076-6879(10)75010-2

Chapter 5
RNA Dynamics in the Control of Circadian Rhythm

Giorgia Benegiamo, Steven A. Brown, and Satchidananda Panda

Abstract The circadian oscillator is based on transcription-translation feedback loops that generate 24 h oscillations in gene expression. Although circadian regulation of mRNA expression at the transcriptional level is one of the most important steps for the generation of circadian rhythms within the cell, multiple lines of evidence point to a disconnect between transcript oscillation and protein oscillation. This can be explained by regulatory RNA-binding proteins acting on the nascent transcripts to modulate their processing, export, translation and degradation rates. In this chapter we will review what is known about the different steps involved in circadian gene expression from transcription initiation to mRNA stability and translation efficiency. The role of ribonucleoprotein particles in the generation of rhythmic gene expression is only starting to be elucidated, but it is likely that they cooperate with the basal transcriptional machinery to help to maintain the precision of the clock under diverse cellular and environmental conditions.

Keywords eRNA • Chromatin modifications • Nascent-seq • RNA-seq • RNAPII • Exon array • IRES • Ribosome • PolyA tail length

G. Benegiamo
Institute of Pharmacology and Toxicology, University of Zürich,
Winterthurerstrasse 190, Zürich 8057, Switzerland

Salk Institute for Biological Studies,
10010, North Torrey Pines Road, La Jolla, CA 92037, USA

S.A. Brown
Institute of Pharmacology and Toxicology, University of Zürich,
Winterthurerstrasse 190, Zürich 8057, Switzerland

S. Panda (✉)
Salk Institute for Biological Studies,
10010, North Torrey Pines Road, La Jolla, CA 92037, USA
e-mail: satchin@salk.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine
and Biology 907, DOI 10.1007/978-3-319-29073-7_5

1 Introduction

To adapt to predictable daily changes in the environment, organisms developed mechanisms to anticipate these changes and respond appropriately. Central to this coordination is an intrinsic *oscillator* that generates *circadian rhythms* of behavior, physiology and metabolism. Anatomically, in mammals, the hypothalamic Suprachiasmatic Nucleus (SCN) consisting of ~20,000 neurons functions is the master circadian oscillator. However, the molecular mechanism of the circadian clock is cell autonomous and is present in almost every cell of our body. The SCN oscillator uses synaptic and diffusible factors to orchestrate rhythms in peripheral tissues in appropriate phases.

The molecular circadian oscillator is based on transcription-translation feedback loops with time-delays that generate 24-h oscillations in many of its constituents. Circadian rhythms in animals are endogenous self-sustaining ~24 h rhythms generated by the basic mechanism of cell autonomous transcriptional feedback loops conserved from flies to human. Both components and mechanisms of circadian rhythms are largely conserved in animals [1]. The transcriptional activators CLOCK and BMAL1 dimerize and activate the transcription of Period (*per*) and Cryptochrome (*cry*). The PER/CRY heterodimer, in turn, represses the transcriptional activities of CLOCK/BMAL1 [2]. In an interconnected loop, nuclear hormone receptors REV-ERB and ROR act as repressors and activators to drive rhythmic transcription of several clock components [3, 4].

The molecular circadian clock drives rhythmic transcription from a large number of genes by (a) directly binding to the respective *cis*-regulatory sites, (b) indirectly by their immediate targets that are also transcription factors, (c) by post transcriptional regulations, and (d) by functional interactions with several signaling and transcriptional regulators. In any given tissue in insects and mammals, up to 15 % of the expressed genome exhibit a circadian expression pattern, with peak levels of expression of different transcripts timed to different times of the day or night [1].

Genomics studies have shown that the steady-state levels of a large number of transcripts show daily oscillations. Immediately after transcription initiation and throughout their life cycles, RNAs are bound by a large number of proteins, some of which remain stably bound while others are subject to dynamic exchange. These complexes containing RNAs and their associated proteins constitute the ribonucleoprotein particles (RNPs). The combination of factors binding to a particular RNA and their position along the transcript determines every step of RNA regulation throughout its lifetime. Given the large number of protein coding, small RNAs, miRNAs, ncRNAs that show circadian rhythm in different tissues, and the indirect evidence for circadian rhythm in ribosome turnover, proteins that bind to these RNAs to process, transport, stabilize, translate or degrade can have a profound impact on circadian rhythms in cellular and organismal function.

Upon transcription, not all mRNAs immediately enter the translationally active pool. Some destined for a particular subcellular location travel in multi-mRNA packets or particles and are held in a quiescent state awaiting either proper subcellular localization or a signal that timing is now right to make protein [5]. Similarly it is

also conceivable that some RNAs accumulate in ribonucleoprotein bodies inside the nucleus waiting to be processed or exported until a second signal is received. This integrated model for the regulation of gene expression fits very well to the features of biological clocks, whose function is to synchronize and adapt internal biological processes to environmental stimuli. Furthermore, the time lag between nascent RNA and mRNA for many of the circadian transcripts is specific to the transcript, which suggests that some type of post-transcriptional regulation at splicing or polyadenylation must play a role to maintain such phase relationship. Posttranscriptional events can also buffer variable transcriptional output to generate robust and reproducible rhythms of mRNA expression and protein synthesis. Overall, there is ample evidence for potential roles of RBPs in circadian organization; however, there are very few RBPs with known function in the circadian function. We will review all the different levels at which gene expression shows circadian variation, from transcription initiation to mRNA stability and translation efficiency and cite specific examples of RBPs regulating circadian function.

2 Circadian Regulation of Transcription Initiation

With increased understanding of the mechanism of transcription initiation, it is becoming apparent that in addition to the RNA polymerase complex, several proteins involved in chromatin modification and several RNAs, including enhancer RNAs, non-coding RNAs, and antisense RNAs, play an integral role in initiation and initiation rate. Transcription is the first step of gene expression, and it starts with the binding of the enzyme RNA polymerase II to a segment of DNA. Of particular importance for transcription initiation is the carboxy-terminal domain (CTD) of RNA polymerase II.

The RNA polymerase II CTD typically consists of up to 52 repeats of the sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser and it serves as a flexible binding scaffold for numerous nuclear factors, determined by the phosphorylation patterns on the CTD repeats. RNA polymerase II can exist in two main forms: RNAPII0, with a highly phosphorylated CTD, and RNAPIIA, with a nonphosphorylated or hypophosphorylated CTD. The phosphorylation state changes as RNAPII progresses through the transcription cycle: the initiating RNAPII is form IIA, and the elongating enzyme is form II0. The phosphorylated CTD physically links pre-mRNA processing to transcription by tethering processing factors to elongating RNAPII, e.g., 5′-end capping, 3′-end cleavage, and polyadenylation. Nearly all of our knowledge of genome-wide transcriptional and post-transcriptional regulation of circadian transcription comes from experiments done in several labs on male mouse liver. In these experiments the adult mice are fed a normal diet ad libitum and entrained to 12 h light: 12 h dark (LD) cycle for a few days. If the mice are held under LD cycle conditions during sample collection, the sampling times are represented as Zeitgeiber time (ZT), where the time of lights-on is considered ZT0. If the mice are transferred to constant darkness prior to sample collection, the timing is represented as Circadian time (CT), where CT0 roughly corresponds to the subjective time of lights-on (or equivalent to ZT0). The

data collected from different times of at least a full day is analyzed by fitting to a wave function, so that the probability values related to robustness of oscillation, peak and trough time/phase of oscillation can be derived. Detected oscillation in given molecules or their activities and the associated phase or time of peak or trough level can be used to explain potential sequence of regulatory events. Due to changes in experimental conditions and sampling frequency, the circadian rhythm parameters might be slightly different in different manuscripts. Therefore, it is relevant to compare the phase information from the same experiment.

Among several published circadian ChIP-seq experiments, the large number of histone modifications, oscillator components binding to chromatin, RNA pol binding, and transcripts from mouse liver reported in Koike *et al*. makes this study relevant to compare the timing of various steps in circadian transcription regulation. In this study CTD phosphorylation status shows a circadian variation, with RNAPIIA signal reaching its peak level at CT14.5 and the RNAPII0 signal peaking at CT0.6. The hypophosphorylated RNAPIIA peak at CT14.5 coincides closely with the peak of intron-containing or nascent transcripts at CT15.1. On the other hand, the peak of hyperphosphorylated RNAPII0 signal coincides with the peak of CRY1 occupancy at CT0.4. At this time CLOCK and BMAL1 are beginning a new cycle of binding but are transcriptionally silent, likely because CRY1 is bound at the same sites. One possible scenario is that RNAPII can be recruited by CLOCK:BMAL and that RNAPII initiation occurs but then pauses or stalls and accumulates at CT0. Alternatively, RNAPII0 occupancy at CT0 could be independent of CLOCK:BMAL1 and could reflect a peak of transcription at CT0 [6].

In addition to circadian oscillations at promoter activity, oscillations in enhancer activities are also found. Fang et al. recently identified circadian transcriptional activity at enhancer regions in mouse liver. This transcriptional activity produces cycling enhancer associated non-coding RNAs called enhancer RNAs (eRNAs; [7, 8]). The circadian eRNAs oscillate in diverse phases and each phase group is enriched in binding motives for different classes of clock transcription factors. The motif enrichment in a given eRNA group is predictive of the specific transcription factor binding. Moreover, circadian eRNAs transcription correlates and can predict rhythmic transcription of nearby genes. The authors propose that circadian transcription factors like CLOCK/BMAL and Rev-erba can bind multiples sites in the genome; however, many of these genes are bound but not controlled, due to inactive binding or long distance looping at different genes. Transcriptional activity at enhancers can be used as markers to assess where a transcription factor is actually functional and they suggest that only the genes that are expressed in phase with CLOCK/BMAL binding are true BMAL1/CLOCK targets [9].

Also chromatin modifications associated with transcription initiation and elongation show circadian variations. Histone H3K4me3, H3K9ac, and H3K27ac are enriched at promoters and show robust circadian rhythms in occupancy at transcription starting sites (TSSs). Histone H3K4me1, a marker that is characteristic of enhancer sites and gene bodies, exhibits a very subtle circadian modulation. There is an antiphase rhythm in H3K4me1 and H3K4me3 occupancy at the Dbp intron1 site. Histone H3K27ac is also highly enriched at both intragenic and intergenic enhancer

sites. The elongation marks, H3K36me3 and H3K79me2, also express very low-amplitude circadian modulation. On a genome-wide level, circadian rhythms in RNAPIIA, RNAPII0, H3K4me3, H3K9ac, and H3K27ac occupancy at TSSs can be seen in all classes of expressed genes and are stronger in intron RNA cycling genes. Unexpectedly, noncycling intron expressed genes also show circadian modulation of RNAPII occupancy and histone modifications. Genome-wide analysis of the periodicity and phase of these histone marks reveals that large number of genes exhibit circadian rhythms in histone modifications. The overall number of histone modification sites does not appear to vary on a circadian basis; rather the recruitment (and initiation) of RNAPII appears to underlie the variation in the amplitude on histone marks. Circadian modulation of RNAPII occupancy and histone modifications occurs not only at promoter proximal regions but also at distal intergenic enhancer sites. Thus, circadian transcriptional regulation appears to be involved in the initial stages of RNAPII recruitment and initiation and the histone modification associated with these events to set the stage for gene expression on a global scale [6].

3 Circadian Regulation of Transcription Termination

Although most focus in circadian regulation has been on daily oscillations in transcription initiation by alternating action of activators and suppressors, recent results have indicated that transcription termination may also be rhythmic. In mammals, PER and CRY proteins accumulate, form a large nuclear complex, and associate with the dimeric transcription factor CLOCK-BMAL at *Per* and *Cry* promoters, repressing their own transcription. PER complexes include the RNA-binding protein NONO and the histone methyltransferase WDR5 [10], and they function in part by recruiting a SIN3 histone deacetylase complex to clock gene promoters [11]. In addition to known PER-associated proteins, Padmanabahn et al. identified the RNA helicases DDX5, DHX9 and SETX in mouse PER complexes. DDX5 and DHX9 function in transcription and pre-mRNA processing [12]. Both are associated with elongating RNA polymerase II [13] and are components of the 3′ transcriptional termination complex [14]. After cleavage of the nascent transcript, unwinding of the RNA-DNA duplex by SETX at the 3′ termination site permits the XRN2 nuclease to degrade the downstream 3′ RNA and thereby release the polymerase [15]. PER complex inhibits SETX activity, blocking subsequent processing by XRN2 and thus blocking transcription termination. Inhibition of termination reduces the rate of transcription, and as a consequence during the negative feedback, RNA polymerase II accumulates at the 5′ site as well as at the 3′ site of *Per* and *Cry* genes [16]. The mammalian PER complex has at least two actions in circadian feedback. It represses transcription by recruiting a SIN3 hystone deacetylase complex to clock gene promoters [11], and it inhibits termination by antagonizing the action of SETX at the 3′ termination site. Both processes contribute to circadian Per gene repression, but one or the other could predominate at different target genes [16].

4 Evidence for the Relevance of Posttranscriptional Events on Circadian Gene Expression

Although circadian regulation of mRNA expression at the transcriptional level is one of the important steps for circadian rhythms in cellular function, multiple lines of evidence point to a disconnect between transcript oscillation and protein oscillation, which can be explained by regulatory RNA-binding proteins acting on the transcripts. The transcription rate itself varies significantly in cells from different tissues [17], yet the free-running endogenous circadian oscillator shows remarkable stability in period length among tissue types. Thus it is likely that the oscillator possesses a mechanism that allows compensating for differences in transcription rates. Indeed, partial inhibition of transcription by a-amanitin treatment in mouse fibroblasts reduces RNA PolII dependent transcription rate by up to threefolds, but doesn't stop cell-autonomous oscillator. The circadian oscillator as measured by translated protein continues to oscillate, albeit with dampened amplitude and slightly shorter periodicity [18]. Similarly, in *Drosophila*, constitutively high mRNA expression of a circadian clock component does not stop the clock; rather, the translated protein level continues to oscillate. These results suggest a posttranscriptional mechanism potentially involving mRNA-binding proteins that can support translational rhythm even when transcriptional oscillation is blunted or inhibited [18, 19].

Furthermore a systematic analysis of the mammalian "circadian proteome" revealed that up to 20 % of soluble proteins in mouse liver are subject to circadian control; however, almost half of the cycling proteins lack a corresponding cycling transcript, further supporting the hypothesis that posttranscriptional mechanisms play a significant role in mammalian circadian rhythms [20, 21]. Interestingly, human red blood cells, which have no nucleus (or DNA) and therefore cannot perform transcription, display robust, temperature-entrainable and temperature-compensated circadian rhythms in peroxiredoxin redox cycles, consistent with the presence of a circadian clock within these cells despite the lack of ability to make new RNA [22].

More recently, additional evidence underlined the importance of posttranscriptional regulation in circadian gene expression. Through high-throughput sequencing or Nascent-Seq and RNA-Seq, Rodriguez at al. identified 136 robust "nascent cyclers" and 237 robust mRNA cyclers in fly heads. Despite a highly significant overlap, most genes in the two data sets are distinct. They propose a model in which transcriptionally active genes can be organized in four groups: genes with both robust nascent and robust mRNA cycling, genes with robust nascent RNA cycling but little or no mRNA cycling, genes with robust mRNA cycling with weak nascent RNA oscillations and genes with robust mRNA cycling but even weaker, and perhaps no, nascent RNA cycling [23].

Menet et al. conducted a similar study in mouse liver. Although many genes are rhythmically transcribed in mouse liver (~15 % of all detected genes), only 42 % of these show mRNA oscillation. More importantly, about 70 % of the genes that exhibit rhythmic mRNA expression do not show transcriptional rhythms, suggesting that posttranscriptional regulation plays a major role in defining the rhythmic mRNA landscape. Also the way in which CLOCK:BMAL1 regulates the transcription of its

target genes differs from what would be expected. Although CLOCK:BMAL target genes are significantly enriched for rhythmic transcribed genes, there is a large discrepancy between the phases of rhythmic BMAL1 DNA binding and those of rhythmic transcription. This is because BMAL1 binding is essentially uniform at ZT3-5, whereas the transcription peaks are much more broadly distributed. The dramatic, genome-wide disconnect between the phases of rhythmic CLOCK:BMAL1 DNA binding and rhythmic target gene transcription suggests that other transcription factors and/or mechanisms collaborate with CLOCK:BMAL1 binding and are critical to determine the phase of clock gene [24].

Another way of assessing the presence of posttranscriptional regulation in circadian gene expression is to sequence the transcriptome and consider the intron signal as a representation of pre-mRNA expression or nascent transcription and the exon signal as a representation of mRNA expression, which can reflect not only transcriptional activity but also posttranscriptional processing events. Koike and collegues quantified intron and exon signals for cycling transcripts in mouse liver and found 1371 intron and 2037 exon RNA cycling transcripts. The intron cycling transcripts are clustered, whereas the exon cycling transcripts have three peak phases. Only 458 genes are in common, and this set of common genes is enriched for known circadian clock genes and high-amplitude cycling target genes reported previously. The phases of the common intron and exon cycling transcripts are correlated, suggesting that transcriptional cycles primarily drive these mRNA rhythms. In the intron-cycling/exon non-cycling class, the cycling pre-mRNA transcripts are clustered at the same phase as the overall intron cycling class, but the steady-state mRNAs are likely to have long half-lives, which would dampen oscillation generated at the transcriptional level. By contrast, in the intron not cycling/exon cycling class, the phases are widely distributed as seen in the overall exon cycling class, and these rhythms likely arise from posttranscriptional regulatory processes such as circadian changes in RNA splicing, polyadenylation, or mRNA stability [6].

5 Circadian Regulation of Alternative Splicing

Alternative splicing is a regulated process during gene expression that results in a single gene coding for multiple proteins. In this process, particular exons of a gene may be included within, or excluded from, the final, processed messenger RNA (mRNA). Alternative splicing is of particular importance amongst the posttranscriptional processes that regulate gene expression, as it allows the human genome to direct the synthesis of many more proteins than would be expected from its 20,000 protein-coding genes.

Alternative splicing is widespread in mammalian genes, affecting approximately 95 % and 80 % of multi-exon genes in humans and mice, respectively [25, 26]. Moreover, alternative splicing is highly regulated by the activity, abundance and binding position of various splicing factors and heterogeneous nuclear ribonucleoproteins, and by the kinetics of transcription elongation and chromatin modifications [27–29].

Using an Affymetrix exon array, McGlincy and collegues demonstrated that the circadian clock regulates alternative splicing in the mouse. 55 exon-probesets from 47 genes where identified to have significant circadian variation. The 47 genes were enriched in pathways representing the circadian clock itself, drug detoxification, caffeine and retinol metabolism and the peroxisome proliferator-activated receptor (PPAR) signaling pathway. The circadian regulation of alternative splicing is tissue dependent, in terms of both phase and amplitude. For some of the exons identified the temporal relationship between alternative splicing and transcript level expression was preserved across tissues, suggesting that these two processes may be coupled in these particular cases. Fasting conditions modulate circadian alternative splicing in an exon dependent manner, but they also modulate the temporal relationship between circadian alternative splicing and circadian mRNA abundance in a gene-dependent manner. Moreover, the alternative splicing of the identified exons is under the control of the local liver clock [30].

At least 15 splicing factors where shown to be robustly cycling in the mouse liver. They include well characterized regulators of alternative splicing (*Srsf3*, *Srsf5*, *Tra2b* and *Khdrbs1*(*Sam68*)), a component of the U2 snRNP (*Sf3b1*), two RNA helicases (*Ddx46* (also component of U2 snRNP) and *Dhx9*), three hnRNP proteins better known for their roles regulating RNA stability and translation (*Hnrnpdl*, *Cirbp* (hnRNP-A18) and *Pcbp2* (hnRNP E2)), and six other proteins with less well characterized roles in RNA processing (*Gtl3*, *Rbms1*, *Thoc3*, *Pcbp4* and *Topors*). Some of these circadian splicing factors were under the control of the local liver clock, while others are likely rhythmic in response to systemic rhythmic cues. Some of the known exon targets of these splicing factors where previously identified to be cycling exons. The discovery of robustly circadian splicing factors, and the fact that a number of their previously characterized target exons are circadian, provide candidates for further study into the molecular mechanisms regulating circadian exons and other posttranscriptional processes [30].

Recently, Preußner et al. demonstrated that the rhythmic alternative splicing of the mRNA encoding U2-auxiliary-factor 26 (U2AF26) contributes to the regulation of Period 1 stability. More in detail they found that U2AF26 undergoes circadian alternative splicing of exons 6 and 7 in peripheral clocks (U2AFΔE67) and that the splicing switch generates a shift in the mRNA reading frame. Skipping of U2AF26 exons 6 and 7 generates a domain with homology to *Drosophila* TIM and enables cytoplasmic, circadian expression of the U2AF26ΔE67 isoform. Furthermore, U2AF26ΔE67 interacts with PER1 and induces its proteasomal degradation; this limits the light induced increase of PER1 and it's proposed as buffering mechanism against sudden light changes [31].

6 Circadian Polyadenylation

The addition of a poly(A) tail to a primary transcript RNA is known as RNA polyadenylation. In nuclear polyadenylation, a poly(A) tail is added to an RNA at the end of transcription. The poly(A) tail consists of multiple adenosine monophosphates; in other words, it is a stretch of RNA that has only adenine bases. In eukaryotes, polyadenylation

is part of the process that produces mature messenger RNA (mRNA) for translation. The poly(A) tail protects the mRNA molecule from enzymatic degradation in the cytoplasm and aids in transcription termination, export of the mRNA from the nucleus, and translation [32]. Almost all eukaryotic mRNAs are polyadenylated [33]. The tail is shortened over time, and, when it is short enough, the mRNA is enzymatically degraded [32]. Regulation of poly(A) tail length is traditionally considered to be unidirectional, going from long to short. However more recent evidence has demonstrated that the ultimate poly(A) tail length is determined by a balance between concomitant deadenylation and polyadenylation, and this balance is controlled in a highly regulated and mRNA-specific manner. In some cases mRNAs with short poly(A) tails can be stored for later activation by re-polyadenylation in the cytosol [34].

Some evidences suggest that poly(A) tail length regulation may take part in controlling circadian-regulated rhythmic gene expression. The deadenylase nocturnin (NOC) removes poly(A) tails from its target RNAs and this process is thought to control target RNA expression by either enhancing RNA degradation or silencing translation. NOC shows rhythmic expression in many tissues such as spleen, kidney and heart in mice with peak levels at the time of light offset. This rhythmicity has been shown to be particularly robust in liver. The mouse NOC gene (*mNoc*) is expressed in a broad range of tissues and in multiple brain regions including suprachiasmatic nucleus and pineal gland. The widespread expression and rhythmicity of *mNoc* mRNA parallels the widespread expression of other circadian clock genes in mammalian tissues, and suggests that NOC plays an important role in clock function or as a circadian clock effector [35]. *mNoc* is also an immediate early gene, and its expression is acutely induced by stimuli such as serum and 12-O-tetradecanoyl-phorbol-13-acetate (TPA) in cultured cells. Remarkably, *mNoc* is the unique deadenylase induced by serum shock. Thus NOC may act in turning off the expression of genes that are required to be silenced as a response to extracellular signals [36].

More recently, it has been shown that 2.3 % of all expressed mRNA exhibit statistically significant rhythmicity in the poly(A) tail length (i.e. the ratio between the "long tail" and the "short tail" fraction of an RNA). The "poly(A) rhythmic" (PAR) mRNAs include mRNAs with peak tail lengths at all phases of the daily cycle but with significantly higher numbers of mRNAs with peak long/short ratios during the night. Based on the pre-mRNA and steady-state mRNA profiles, the PAR mRNAs can be categorized into three classes: Class I PAR mRNAs (49.2 %) are rhythmic in their poly(A) tail length and pre-mRNA and steady-state mRNA levels, class II PAR mRNAs (32.3 %) are rhythmic in their poly(A) tail length and pre-mRNA expression but not in steady-state mRNA levels, and class III PAR mRNAs (18.5 %) are rhythmic in their poly(A) tail length rhythms but not in pre-RNA or steady-state mRNA levels. There are significant differences in mRNA half-lives among the different PAR classes; class III mRNAs are the most stable, followed by class II mRNAs, with class I mRNAs being the least stable [37].

The rhythmic poly(A) lengths of both class I and II mRNAs reflect nuclear polyadenylation (likely by the canonical poly(A) polymerase α), coordinated with rhythmic transcription during classical 3′ end processing. The defining characteristics of these two classes are the differences in the steady-state mRNA rhythmicity and mRNA

stability, suggesting that the lack of rhythmicity in the class II PARs reflects the longer half-lives of these mRNAs. The mechanism of poly(A) rhythmicity in both class I and II mRNAs results from the addition of long tails following rhythmic synthesis and subsequent deadenylation that does not cause immediate decay. This delay in decay is more pronounced in the class II mRNAs and results in the arrhythmic steady-state levels. It appears that class I/II PAR mRNAs can exist in short-tailed states and that rhythmic control of poly(A) tail length is somehow correlated with delayed accumulation of steady-state mRNA and may be part of a regulatory mechanism to regulate the timing of mRNA/protein rhythmicity [37].

Class III mRNAs exhibit robust rhythmicity in their poly(A) tail length, yet are not rhythmically transcribed and have longer half-lives. Thus, class III PARs must employ transcription-independent mechanisms to control their rhythmic poly(A) tail lengths. Peak distribution analysis of class III PAR mRNAs revealed that >80 % had their longest poly(A) tails during the day, which is distinct from the nighttime poly(A) rhythmic profile of class I and many of the class II mRNAs. The poly(A) rhythms of class III mRNAs are likely to be controlled by rhythmic cytoplasmic polyadenylation. Indeed, the steady-state mRNA level of several putative cytoplasmic polyadenylation machinery components in the liver, including *Cpeb2*, *Cpeb4*, *Parn*, and *Gld2*, are rhythmically expressed with phases similar to the majority of the class III PAR mRNAs, peaking in the early day [37].

In mouse liver rhythmic poly(A) tail lengths correlate strongly with the ultimate circadian protein expression profiles, with the protein peaking ~4–8 h after the time of the longest poly(A) tail. Therefore poly(A) tail rhythms can generate rhythmic protein levels even when there is no rhythm in the steady-state mRNA levels [37].

7 Regulation at Translation Initiation and Ribosome Biogenesis

Many oscillating proteins in the mouse liver are encoded by constantly expressed mRNAs and among the rhythmically expressed genes in the liver, there are several genes encoding proteins involved in mRNA translation, including components of the translation pre-initiation complex [38]. The circadian clock controls the transcription of translation initiation factors as well as the rhythmic activation of signaling pathways involved in their regulation. Initiation in eukaryotes requires at least ten proteins, which are designated eIFs (eukaryotic initiation factors). The mRNAs of most of the factors involved in translation initiation are rhythmically expressed with a period of 24 h. There isn't a significant variation in protein abundance, but these factors undergo strong rhythmic phosphorylation [38]. The initiation factors eIF4E and eIF4G, in association with eIF-4A and eIF-4B, are involved in binding the mRNA and bringing it to the 40S ribosomal subunit. eIF4E, which recognizes the 5′ cap of the mRNA, is mostly phosphorylated during the day, with a peak at the

end of the light period (ZT6-12). eIF4G, eIF4B and 4E-BP, and ribosomal protein (RP) S6 (RPS6) are mainly phosphorylated during the night, which is, in the case of nocturnal animals like rodents, the period when the animals are active and consume food. Phosphorylation of these factors is well characterized and involves different signaling pathways whose reported activity perfectly correlates with the observed phosphorylation rhythm [38]. eIF4E is phosphorylated by the extracellular signal-regulated protein kinase (ERK)/mitogen activated protein kinase (MAPK)-interacting kinase (MNK) pathway, which is most active during the day, at the time when eIF4E reaches its maximum phosphorylation. On the other hand, eIF4G, eIF4B, 4E-BP1, and RPS6 are mainly phosphorylated by the target of rapamycin (TOR) complex 1 (TORC1), which is activated during the night at the time when the phosphorylation of these proteins reaches its maximum level. It has been shown that mTOR, its partner Raptor, as well as its regulating kinase Map3k4, are also rhythmically expressed, thus potentially further contributing to the rhythmic activation of TORC1. The rhythmic phosphorylation of 4E-BP1 results in the release eIF4F, allowing its binding to the mRNA and the initiation of translation [38].

The polysomal RNA fraction (RNA sub-fraction composed mainly of actively translated RNAs) in mouse liver also follows a diurnal cycle, showing that a rhythmic translation does occur in this tissue. Approximately 2 % of the expressed genes are translated with a rhythm that is not explained by rhythmic mRNA abundance as in most cases the total mRNA levels are constant. Among translationally regulated genes, 70 % were found in the polysomal fraction during the same time interval, starting at ZT8 before the onset of the mouse feeding period and finishing at the end of the dark period [38].

The circadian clock was shown to regulate also ribosome biogenesis by influencing the transcription of ribosomal protein (RP) mRNAs and ribosomal RNAs (rRNAs). RPs show a rhythmic abundance with highest expression during the night. Pre-mRNA accumulation of several RPs exhibits a rhythmic pattern too, with a peak at ZT8, just before the activation of their translation. As for rRNA transcription, the synthesis of the ribosome constituent precursor 45S rRNA (containing 28S, 5.8S and 18S rRNAs) is rhythmic and synchronized with RP mRNAs transcription, indicating that all elements involved in ribosome biogenesis are transcribed in concert and coordinated with the feeding period. In mammals rRNA transcription is highly regulated by the upstream binding factor (UBF). Not surprisingly UBF1 is rhythmically expressed in mouse liver too, at both mRNA and protein levels, in phase with RP mRNAs and rRNA transcription. Mice devoid of a functional circadian clock lose the rhythmic activation of TORC1 and ERK signaling pathways, and the rhythmic expression of UBF1. In addition these animals show lack of synchrony and coordination of 45S rRNA and RP pre-mRNA transcription, highlighting the crucial role of the circadian clock in this mechanism. Ribosomal protein synthesis in eukaryotes is a major metabolic activity that involves hundreds of individual reactions; this energy-consuming process has to be confined to a time when energy and nutrients are available in sufficient amounts, which, in the case of rodents, is during the night [38].

8 RNA-binding Proteins Regulating mRNA Stability and Translational Efficiency Are Important for Oscillation of Core Clock Components

The number of RNA-binding proteins and those with RNA-binding motifs encoded by the human genome is remarkable. As an example a single type of RNA-binding domain, the RNA recognition motif (RRM), is represented in nearly 500 different human genes. RNA-binding proteins (RBPs) couple transcription and subsequent post-transcriptional steps by interacting with their target transcripts. Even though some RBPs bind to common elements present in almost every mRNA in a sequence-independent and nonspecific manner, the majority of RNA-binding factors target particular structures or sequences present in some RNAs but not others [39]. The posttranscriptional events involving multiple mRNAs must be highly coordinated and RBPs, including export proteins, provide coordinating functions at all steps along the posttranscriptional regulatory chains. RBPs allow the mRNA molecules to interface with other intracellular machineries mediating their splicing, transport, stabilization or degradation, localization, or translation into protein, as well as the response to stimuli. Indeed individual mRNAs contain binding sites for different RBPs and can respond to a wide range on inputs, so that their expression can be adjusted to changing environmental conditions. RBPs are thus the leading actors of an intricate regulatory network, which is equally complex as that controlling initial RNA synthesis. Because RBPs can bind to more than one RNA with sequence specificity, the existence of a "posttranscriptional operon" has been postulated whose function is to expand the regulatory plasticity of our relatively "small" genome. In fact the expression of proteins with common functional themes or subcellular distributions is coordinated by large-scale regulatory networks operating at the mRNP level. The final outcome of protein synthesis is thus an mRNP-driven process that responds dynamically to the environment and cellular growth conditions [5, 39].

In particular the posttranscriptional regulation of mRNA stability and translational efficiency are often mediated by cis elements in mRNAs that interact with RNA-binding proteins and/or microRNAs. In most cases, these cis elements reside in the 3′ untranslated region (UTR), and several 3′ UTR motifs have been identified that are critical for mRNA splicing, transport, stability, localization, and translation.

The 3′UTR-dependent mRNA decay is involved in the regulation of circadian oscillation of Period 2 (*per2*) mRNA. In particular the polypyrimidine tract-binding protein (PTB), also known as heterogeneous nuclear ribonucleoprotein I (hnRNP), binds to *per2* 3′UTR and has an mRNA destabilizing activity. Indeed the cytoplasmic PTB expression pattern is reciprocal with *per2* mRNA oscillation and depletion of PTB with RNAi results in *per2* mRNA stabilization [40]. A similar study reported that the 3′UTR is also important for the mRNA stability of another core clock component, mouse cryptochrome 1 (*cry1*). The 3′UTR of *cry1* contains a destabilizing cis-acting element that contributes to the stability of *cry1* mRNA. The binding of hnRNP D to *cry1* 3′UTR is responsible for the rapid decay of *cry1* mRNA during its declining phase and modulates *cry1* circadian rhythm [41].

Also, the stability of mouse Period3 (*per3*) is dramatically changed in a circadian phase-dependent manner. In the case of (*per3*), the control of its circadian mRNA stability requires the cooperative function of both the 5′ and 3′ UTRs. Several studies reported that mRNA stability can be regulated by the 5′UTR, a mechanism called translational regulation-coupled mRNA decay. In such cases translational inhibition causes mRNA stabilization. Similarly hnRNP Q binds to both 5′ and 3′UTR in the *per3* mRNA and not only reduces the translation efficiency but also increases the mRNA stability. *per3* mRNA decay is connected to its translation kinetics and the central region of *per3* 5′UTR is responsible for coupling of translation and mRNA decay. The binding of hnRNP Q of *per3* 5′UTR is phase dependent and maintains robust mRNA oscillation [42].

Mouse LARK, another RBP, has been shown to activate the posttranscriptional expression of the mouse period1 (*per1*) mRNA. A strong circadian cycling of the LARK protein is observed in the suprachiasmatic nuclei with a phase similar to that of PER1, although the level of the *lark* transcripts are not rhythmic. LARK protein binds directly to a cis-element in the 3′ UTR of the *per1* mRNA and causes increased PER1 protein levels, by activating *per1* mRNA translation. Alterations of *lark* expression in cycling cells causes significant changes in circadian period, with *lark* knockdown by siRNA resulting in a shorter circadian period, and *lark* overexpression resulting in a lengthened period [43].

Many studies have shown that mammalian cells utilize internal ribosome entry site (IRES)-mediated translation for rapid adaptation to certain environments, such as chemotoxic stress [44], mitosis [45] and apoptosis [46], and generally under conditions when cap-dependent translation is compromised. For IRES-mediated translation, proteins known as IRES *trans*-acting factors (ITAFs) must recognize IRES elements in a structure or sequence-dependent manner. Recent evidences suggest that IRES-mediated translation might be one of the mechanisms regulating the protein oscillation of key clock components like Rev-erb α. Also known as Nr1d1, Rev-erb α was identified as a regulator of lipid metabolism [47]. It also plays an important role in the maintenance of circadian timing in brain and liver tissue [48, 49] and it is a well-known transcriptional repressor in the positive limb of circadian transcription [3, 50]. Kim and colleagues have demonstrated that hnRNP Q and PTB modulate mRev-erb α IRES-mediated translation. Knockdown of hnRNP Q and PTB leads to the alteration of the mRNA levels of several clock genes, thus posttranscriptional regulation by hnRNP Q and PTB is necessary to maintain the circadian feedback loop [7, 8].

9 Conclusion

There is increasing evidence that the RBPs play an important role in homeostatic control of the periodicity of circadian oscillator and its output regulation. RBP mediated regulation of various steps in the transcription, RNA processing and RNA half-life helps maintain the precision of the clock under diverse cellular and environmental conditions. Circadian regulation of cellular physiology and metabolism is mediated by

daily oscillations in the steadystate levels of nascent RNA, mRNA, and protein levels. However, large fraction of oscillating proteins or mRNA does not exhibit a correlated rhythm in nascent RNAs, which suggests that post-transcriptional regulation involving RBPs is most likely involved. Since many RBPs bind and regulate the location, transport, translation of a large number of target RNAs, a rhythmic level of a given RBP likely helps temporally coordinate the function of the target RNAs. Another challenge in circadian regulation is the newly recognized role of eating pattern and nutrition quality in the daily oscillations of RNA and proteins. In rodent liver, the circadian transcriptome in peripheral organs appears to be heavily determined by the nutrition quality and time of eating. This implies that the nutrition information encoded in several metabolites might affect the circadian transcriptome by both transcriptional and post-transcriptional mechanisms and RBPs will likely play an important role in integrating nutrition status with the endogenous circadian oscillator function.

Much of the evidence for the roles of RBPs in circadian regulation is indirect. Although genome-wide transcriptome studies have shown circadian rhythms in the mRNA levels of several RBPs, whether the RBP proteins and their cellular localization are also circadian is yet to be determined. Similarly, as the target RNAs for many of the RBPs are discovered, informatics approaches to integrate these findings with circadian transcriptome datasets will begin to explain post-transcriptional mechanisms of circadian regulation. Overall, the area of investigation on how RBPs are involved in the circadian regulation is a nascent field with plenty of opportunities for discoveries and mechanistic insight.

References

1. Bell-Pedersen D, Cassone VM, Earnest DJ, Golden SS, Hardin PE, Thomas TL, Zoran MJ (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 6:544–556
2. Lee C, Etchegaray JP, Cagampang FR, Loudon AS, Reppert SM (2001) Posttranslational mechanisms regulate the mammalian circadian clock. Cell 107(7):855–867
3. Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U (2002) The orphan nuclear receptor REV-ERBalpha controls circadian transcription within the positive limb of the mammalian circadian oscillator. Cell 110(2):251–260
4. Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, Hogenesch JB (2004) A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. Neuron 43(4):527–537
5. Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science 309(5740):1514–1518
6. Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. Science 338(6105):349–354
7. Kim DY, Woo KC, Lee KH, Kim TD, Kim KT (2010) hnRNP Q and PTB modulate the circadian oscillation of mouse Rev-erb alpha via IRES-mediated translation. Nucleic Acids Res 38(20):7068–7078
8. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF,

Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465(7295):182–187
9. Fang B, Everett LJ, Jager J, Briggs E, Armour SM, Feng D, Roy A, Gerhart-Hines Z, Sun Z, Lazar MA (2014) Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell 159(5):1140–1152
10. Brown SA, Ripperger J, Kadener S, Fleury-Olela F, Vilbois F, Rosbash M, Schibler U (2005) PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. Science 308(5722):693–696
11. Duong HA, Robles MS, Knutti D, Weitz CJ (2011) A molecular mechanism for circadian clock negative feedback. Science 332(6036):1436–1439
12. Fuller-Pace FV (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Res 34(15):4206–4215
13. Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol Cell 26(6):867–881
14. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL (2009) Molecular architecture of the human pre-mRNA 3′ processing complex. Mol Cell 33(3):365–376. doi:10.1016/j.molcel.2008.12.028
15. Skourti-Stathaki K, Proudfoot NJ, Gromak N (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell 42(6):794–805. doi:10.1016/j.molcel.2011.04.026
16. Padmanabhan K, Robles MS, Westerling T, Weitz CJ (2012) Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. Science 337(6094):599–602
17. Schmidt EE, Schibler U (1995) Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. J Cell Biol 128(4):467–483
18. Dibner C, Sage D, Unser M, Bauer C, d'Eysmond T, Naef F, Schibler U (2009) Circadian gene expression is resilient to large fluctuations in overall transcription rates. EMBO J 28(2):123–134
19. So WV, Rosbash M (1997) Post-transcriptional regulation contributes to Drosophila clock gene mRNA cycling. EMBO J 16(23):7146–7155
20. Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, Wong GK, Chesham J, Odell M, Lilley KS, Kyriacou CP, Hastings MH (2006) Circadian orchestration of the hepatic proteome. Curr Biol 16(11):1107–1115
21. Robles MS, Cox J, Mann M (2014) In-Vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet 10(1), e1004047
22. O'Neill JS, Reddy AB (2011) Circadian clocks in human red blood cells. Nature 469(7331): 498–503
23. Rodriguez J, Tang CH, Khodor YL, Vodala S, Menet JS, Rosbash M (2013) Nascent-Seq analysis of Drosophila cycling gene expression. Proc Natl Acad Sci U S A 110(4):E275–E284
24. Menet JS, Rodriguez J, Abruzzi KC, Rosbash M (2012) Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. Elife 1, e00011
25. Mollet IG, Ben-Dov C, Felício-Silva D, Grosso AR, Eleutério P, Alves R, Staller R, Silva TS, Carmo-Fonseca M (2010) Unconstrained mining of transcript data reveals increased alternative splicing complexity in the human transcriptome. Nucleic Acids Res 38:4740–4754
26. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476
27. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. Cell 144:16–26
28. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463
29. Witten JT, Ule J (2011) Understanding splicing regulation through RNA splicing maps. Trends Genet 27:89–97
30. McGlincy NJ, Valomon A, Chesham JE, Maywood ES, Hastings MH, Ule J (2012) Regulation of alternative splicing by the circadian clock and food related cues. Genome Biol 13(6):R54

31. Preußner M, Wilhelmi I, Schultz AS, Finkernagel F, Michel M, Möröy T, Heyd F (2014) Rhythmic U2af26 alternative splicing controls PERIOD1 stability and the circadian clock in mice. Mol Cell 54(4):651–662
32. Guhaniyogi J, Brewer G (2001) Regulation of mRNA stability in mammalian cells. Gene 265(1–2):11–23
33. Hunt AG, Xu R, Addepalli B, Rao S, Forbes KP, Meeks LR, Xing D, Mo M, Zhao H, Bandyopadhyay A, Dampanaboina L, Marion A, Von Lanken C, Li QQ (2008) Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. BMC Genomics 9:220
34. Richter JD (1999) Cytoplasmic polyadenylation in development and beyond. Microbiol Mol Biol Rev 63(2):446–456
35. Wang Y, Osterbur DL, Megaw PL, Tosini G, Fukuhara C, Green CB, Besharse JC (2001) Rhythmic expression of Nocturnin mRNA in multiple tissues of the mouse. BMC Dev Biol 1:9
36. Garbarino-Pico E, Niu S, Rollag MD, Strayer CA, Besharse JC, Green CB (2007) Immediate early response of the circadian polyA ribonuclease nocturnin to two extracellular stimuli. RNA 13(5):745–755
37. Kojima S, Sher-Chen EL, Green CB (2012) Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. Genes Dev 26(24):2724–2736
38. Jouffe C, Cretenet G, Symul L, Martin E, Atger F, Naef F, Gachon F (2013) The circadian clock coordinates ribosome biogenesis. PLoS Biol 11(1), e1001455
39. Keene JD (2010) Minireview: global regulation and dynamics of ribonucleic acid. Endocrinology 151(4):1391–1397
40. Woo KC, Kim TD, Lee KH, Kim DY, Kim W, Lee KY, Kim KT (2009) Mouse period 2 mRNA circadian oscillation is modulated by PTB-mediated rhythmic mRNA degradation. Nucleic Acids Res 37(1):26–37
41. Woo KC, Ha DC, Lee KH, Kim DY, Kim TD, Kim KT (2010) Circadian amplitude of cryptochrome 1 is modulated by mRNA stability regulation via cytoplasmic hnRNP D oscillation. Mol Cell Biol 30(1):197–205
42. Kim DY, Kwak E, Kim SH, Lee KH, Woo KC, Kim KT (2011) hnRNP Q mediates a phase-dependent translation-coupled mRNA decay of mouse Period3. Nucleic Acids Res 39(20): 8901–8914
43. Kojima S, Matsumoto K, Hirose M, Shimada M, Nagano M, Shigeyoshi Y, Hoshino S, Ui-Tei K, Saigo K, Green CB, Sakaki Y, Tei H (2007) LARK activates posttranscriptional expression of an essential mammalian clock protein, PERIOD1. Proc Natl Acad Sci U S A 104(6):1859–1864
44. Dobbyn HC, Hill K, Hamilton TL, Spriggs KA, Pickering BM, Coldwell MJ, de Moor CH, Bushell M, Willis AE (2008) Regulation of BAG-1 IRES-mediated translation following chemotoxic stress. Oncogene 27(8):1167–1174
45. Schepens B, Tinton SA, Bruynooghe Y, Parthoens E, Haegman M, Beyaert R, Cornelis S (2007) A role for hnRNP C1/C2 and Unr in internal initiation of translation during mitosis. EMBO J 26(1):158–169
46. Bushell M, Stoneley M, Kong YW, Hamilton TL, Spriggs KA, Dobbyn HC, Qin X, Sarnow P, Willis AE (2006) Polypyrimidine tract binding protein regulates IRES-mediated gene expression during apoptosis. Mol Cell 23(3):401–412
47. Coste H, Rodríguez JC (2002) Orphan nuclear hormone receptor Rev-erbalpha regulates the human apolipoprotein CIII promoter. J Biol Chem 277(30):27120–27129
48. Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell 93(6):929–937
49. Teboul M, Guillaumond F, Gréchez-Cassiau A, Delaunay F (2008) The nuclear hormone receptor family round the clock. Mol Endocrinol 22(12):2573–2582
50. Ramakrishnan SN, Muscat GE (2006) The orphan Rev-erb nuclear receptors: a link between metabolism, circadian rhythm and inflammation? Nucl Recept Signal 4, e009

Chapter 6
Roles of RNA-binding Proteins and Post-transcriptional Regulation in Driving Male Germ Cell Development in the Mouse

Donny D. Licatalosi

Abstract Tissue development and homeostasis are dependent on highly regulated gene expression programs in which cell-specific combinations of regulatory factors determine which genes are expressed and the post-transcriptional fate of the resulting RNA transcripts. Post-transcriptional regulation of gene expression by RNA-binding proteins has critical roles in tissue development—allowing individual genes to generate multiple RNA and protein products, and the timing, location, and abundance of protein synthesis to be finely controlled. Extensive post-transcriptional regulation occurs during mammalian gametogenesis, including high levels of alternative mRNA expression, stage-specific expression of mRNA variants, broad translational repression, and stage-specific activation of mRNA translation. In this chapter, an overview of the roles of RNA-binding proteins and the importance of post-transcriptional regulation in male germ cell development in the mouse is presented.

Keywords Post-transcriptional regulation • Alternative mRNA processing • Splicing • Polyadenylation • Translational control • RNA-binding proteins • Gametogenesis

1 Introduction

Post-transcriptional regulation of gene expression is a central and widespread mechanism that alters genetic output during cell differentiation and tissue development. Transcriptional controls define which combinations of genes are transcribed in a cell and the magnitude of each gene's transcriptional output, while post-transcriptional controls include regulatory events that act on the RNA products of transcription. These include regulation of nuclear pre-mRNA splicing and polyadenylation, and cytoplasmic mRNA localization, stability, translation, and degradation. In different

D.D. Licatalosi (✉)
Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH 44106, USA
e-mail: ddl33@case.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_6

cell types and stages of development, specific combinations of RNA-binding proteins (RBPs) establish and modulate post-transcriptional gene regulatory networks. The importance of RBPs in tissue homeostasis and function is highlighted by a growing list of human diseases associated with aberrant expression of RBPs and disruption of post-transcriptional regulatory events [1].

Alternative mRNA processing (the production of alternatively spliced and polyadenylated mRNAs from a single gene) has major roles in transcriptome diversification in different tissues [2]. Tissue-specific differences also exist in the extent to which mRNAs are translationally controlled. The mammalian testis stands out from other tissues with respect to transcriptome complexity and widespread post-transcriptional regulation [3–5]. Within the testis are seminiferous tubules where germ cells proceed through a well-characterized series of developmental steps to generate haploid gametes. Mouse models have identified essential roles for RBPs and post-transcriptional regulation in nearly all steps of germ cell development, from the earliest embryonic stages to the formation and release of mature spermatozoa [6].

In this chapter, the importance of RBPs and post-transcriptional regulation of protein coding genes in gametogenesis is reviewed, with a focus on male germ cell development in the mouse. As mouse and human male germ cell development are broadly comparable, studies of germ cell development in the mouse have provided insights on many different aspects of human cellular and reproductive biology [7, 8]. This includes new insights into regulatory mechanisms that control mammalian cell differentiation, apoptosis, chromosome biology, infertility, and testicular cancer. In the first section, an overview of the different stages of male germ cell development is provided. In the second section, different types of post-transcriptional regulation and examples of their impact on germ cell gene expression are presented. In the final section, evidence of the importance of specific RBPs and their roles in germ cell development is reviewed.

2 Male Germ Cell Development

2.1 Overview of Pathway

Germ cell development can be broadly divided into distinct stages and involves a number of well-characterized cellular division and differentiation events driven by intrinsic factors and extrinsic cues [9] (Fig. 6.1). The embryonic stage of germ cell

Fig. 6.1 (continued) to a male or female program of development. (**b**) During the first postnatal stage of germ cell development, spermatogonia type A cells self renew or undergo a series of proliferative and differentiating divisions to generate chains of cells that will enter meiosis. (**c**) Meiosis involves a single genome duplication event followed by two successive divisions to generate haploid spermatids. (**d**) Spermiogenesis, the process by which round spermatids progress through 16 steps to transform into spermatozoa that are released into the lumen of the seminiferous tubule

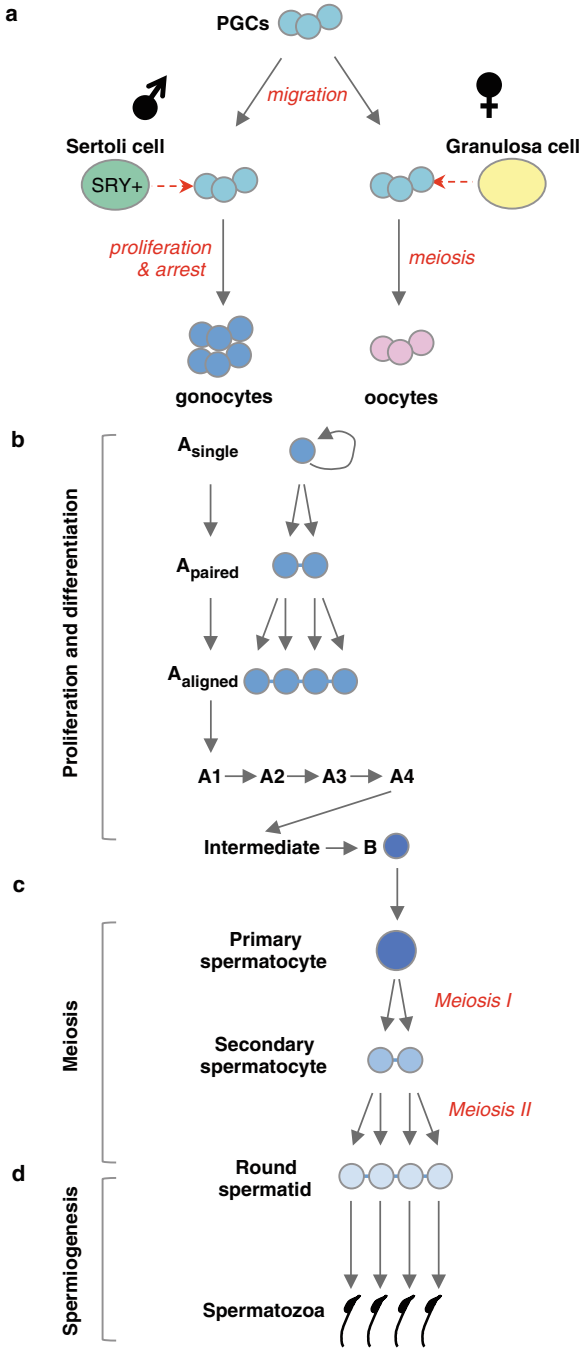


Fig. 6.1 Overview of male germ cell development in the mouse. (**a**) During embryogenesis, primordial germ cells (PGCs) migrate to the genital ridge where they receive signals (*red dashed arrow*) from gonadal support cells (Sertoli cells or granulosa cells in XY and XX embryos, respectively) and commit

development includes the specification (or 'birth') of germ-line cells, their migration to the genital ridge (the site of the future gonads), and sex determination. The remaining stages of germ cell development occur postnatally and include: (1) a proliferative stage where spermatogonial cells either self-renew or undergo a number of proliferative divisions to yield spermatoctyes that enter meiosis; (2) a meiotic stage where four genetically distinct haploid cells (spermatids) are generated from each spermatocyte; and (3) a differentiation stage (called spermiogenesis) where haploid spermatids transform into spermatozoa. Here, an overview of the major stages of male germ cell development in the mouse, with an emphasis on steps in which RBPs and post-transcriptional control of gene expression have critical roles is described.

2.2 Embryonic Stages of Germ Cell Development

Germ-line cells are first detectable at approximately day 7 of embryogenesis (E7) as a cluster of cells in the epiblast [10]. These primordial germ cells (PGCs) proliferate and migrate to the genital ridge, during which time they remain sexually bipotent, able to commit to either the male or female program of germ cell development. Germ cell sex-determination depends on whether gonadal support cells express the SRY gene encoded on chromosome Y. In the absence of SRY expression (XX embryos), the supporting gonadal cells differentiate into female granulosa cells. In XY embryos, SRY expression induces differentiation of gonadal support cells into male Sertoli cells. Sex-specific gonadal support cells (granulosa cells or Sertoli cells) provide extracellular signals that determine whether PGCs progress towards a female program of development and differentiate to meiotic oocytes, or commit to a male program of germ cell development in which PGCs become gonocytes (also called prospermatogonia) and proliferate briefly before undergoing cell cycle arrest at G1/G0 and quiescence for the remainder of embryogenesis (Fig. 6.1a) [11, 12]. In the postnatal testis, Sertoli cells remain in close contact with germ cells, providing structural and nutritional support throughout spermatogenesis via testis-specific junctions [13].

2.3 Postnatal Germ Cell Development

2.3.1 Spermatogonia Proliferation, Renewal, and Differentiation

A few days after birth, the quiescent gonocytes resume mitotic proliferation and differentiate into 'type A' spermatogonial cells (Fig. 6.1b) [14–16]. Single type A spermatogonia (A_{single} or A_s) are thought to have stem cell potential and undergo self-renewing divisions to produce two new A_s cells. Alternatively, A_s cells can divide to produce a pair of spermatogonial cells (A_{paired} or A_{pr}) that remain

connected to one another via intercellular bridges resulting from incomplete cytokinesis. Subsequent divisions yield chains of 4–16, and even 32 spermatogonial cells called $A_{aligned}$ (A_{al}). The differentiation of A_{al} cells into A1 spermatogonia begins the strictly time-regulated stages of spermatogenesis involving mitotic divisions to generate interconnected chains of A2, A3, A4, Intermediate, and finally type B spermatogonia. B spermatogonia undergo a final mitotic division to yield pre-leptotene spermatocytes that enter meiosis.

2.3.2 Meiosis

The second major phase of postnatal germ cell development is meiosis, in which spermatocytes undergo a single genome duplication event followed by two successive divisions (meiosis I and meiosis II) to generate haploid cells called spermatids (Fig. 6.1c). During the prolonged prophase of meiosis I, chromosomes condense and homologous pairs of chromosomes recognize one another and align. The recognition, juxtaposition, and synapsis of homologous chromosomes allows them to physically exchange genetic information through the repair of double strand breaks [17]. Recombination between homologous sequences on maternally- and parentally-inherited chromosomes creates new combinations of alleles and therefore generates genetic diversity. Additional diversity arises from the independent assortment (segregation) of homologous pairs into daughter cells during meiosis I. Meiosis II is significantly shorter than meiosis I and is similar to mitosis in that sister chromatids are separated from one another and segregate into daughter cells.

2.3.3 Spermiogenesis (Spermatid Differentiation)

Round spermatids (the haploid products of meiosis) undergo an ordered series of cytological and morphological changes to produce spermatozoa—slender,) elongated cells that consist of three main regions: (1) a flagellum for motility, (2) a midpiece region lined with mitochondria to provide ATP for motility, and (3) a head region consisting of a compact nucleus whose anterior is encased with a granular vesicle (the acrosome) that contains hydrolytic enzymes necessary for oocyte penetration (Fig. 6.1d) [18].

The process of spermatid differentiation (called spermiogenesis) takes ~13.5 days to complete in the mouse, and consists of 16 steps that can be roughly divided into the round spermatid steps (steps 1–8) and elongating spermatid steps (steps 9–16). During spermatid elongation, extensive chromatin remodeling and compaction results from the successive replacement of histones with transition proteins, followed by the replacement of transition proteins with protamines. A consequence of chromatin compaction is transcriptional inactivation in elongating spermatids [19, 20]. As a result, synthesis of new proteins during spermatid elongation is dependent on a reservoir of translationally-repressed mRNAs synthesized days earlier in transcriptionally active round spermatids (discussed in greater detail below)

[3]. In the final stages of spermiogenesis, mature spermatids shed excess cytoplasm, detach from Sertoli cells as cell junctions are severed, and are released into the lumen of the seminiferous tubule [21]. Mature spermatids (called spermatozoa) then transit out of the testis and into the epididymis where further maturation occurs.

3 Complexity and Post-transcriptional Regulation of the Developing Germ Cell Transcriptome

3.1 Modulating Gene Output Via Alternative Splicing and Polyadenylation

Nearly all protein-coding genes in higher eukaryotes have a 'split-gene' organization in which the sequences present in the resulting mRNA (expressed sequences or exons) are interrupted in the gene by longer intervening sequences (introns) [2, 22]. Consequently, production of a mature mRNA template for translation requires the removal of intronic sequences from the mRNA precursor (pre-mRNA) and the precise joining (or splicing) of exons. Alternative splicing is the process in which specific exons are differentially spliced into the mature transcript. Nearly all multi-exon genes in mammals yield alternatively spliced mRNA isoforms, most of which are expressed in a specific tissue or stage of development [23–26].

In some cases, alternative splicing results in modest changes in the primary sequence and functional properties of the encoded protein, while in others alternative splicing can have a profound effect on biology and act as a switch that controls the production of protein isoforms with antagonist activities (Fig. 6.2). For example, several genes encoding regulators of apoptosis can yield both anti- and pro-apoptotic isoforms as a result of alternative pre-mRNA splicing [27]. Tissue-restricted alternative splicing events frequently alter regions of proteins that are phosphorylated thus altering the range of targets for specific kinases in each tissue [26, 28], as well as regions that specify tissue-specific protein-protein interaction networks [29, 30]. Regulated alternative splicing can also be coupled to changes in mRNA abundance as the inclusion or exclusion of some alternative exons results in a coding frame-shift and the introduction of a premature termination codon which then targets the mRNA for degradation by the nonsense-mediated mRNA decay pathway [31].

With few exceptions, all mRNAs receive a polyadenosine tract (polyA tail) at the 3′end [32]. The addition of the polyA tail is functionally linked to transcription termination and involves two tightly coupled steps [33]. In the first step of 3′end formation, pre-mRNA is endonucleolytically cleaved to expose a free 3′ hydroxyl that will be the substrate for the second step, the non-templated addition of a polyA tail. The majority of mammalian genes yield alternative mRNA variants that can be cleaved and polyadenylated at one of multiple positions [34]. The most common alternative polyA site variants arise from selection of one of multiple polyA sites that are present in tandem in the same exon (Fig. 6.2). In these alternative mRNAs,

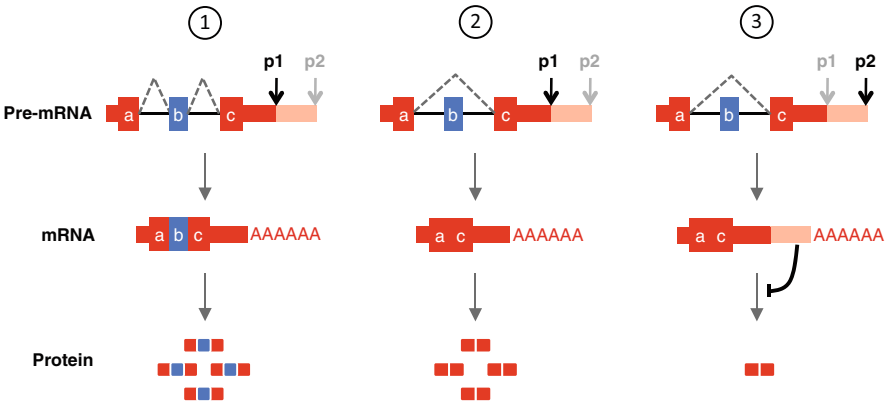


Fig. 6.2 Gene regulation through alternative mRNA regulation. In this example, a single gene yields three identical pre-mRNAs that are alternatively processed into different mRNAs to alter the identity and abundance of the encoded protein. In panels (*1*) and (*2*), alternative splicing of exon 'b' yields mRNAs that encode alternative protein variants. In panels (*2*) and (*3*), alternative polyadenylation yields mRNAs that differ with respect to 3′UTR length, with the long 3′UTR variant (panel (*3*)) possessing regulatory elements that lead to reduced accumulation of the encoded protein

selection of a proximal or distal site for polyA tail addition does not alter protein coding sequences, but does alter 3′UTR length. Thus, alternative polyadenylation has the potential to switch mRNAs from one cytoplasmic fate to another, by altering the repertoire of 3′UTR *cis* regulatory sequences associated with post-transcriptional mRNA control including regulation by small RNAs (miRNAs and possibly piRNAs [35]) or RBPs that control mRNA localization, translation, and decay (Fig. 6.2) [36–39]. Alternatively polyadenylated mRNAs can also arise from differential use of polyA sites located in different 3′ terminal exons. This form of alternative polyadenylation is coupled to changes in exon splicing of the pre-mRNA and results in mRNAs that have distinct 3′UTRs and code for proteins with different C-termini. Thus alternative polyadenylation coupled to alternative splicing can yield alternative mRNAs from the same gene yet code for different protein isoforms and subject to different post-transcriptional controls.

Similar to splice variants, most alternatively polyadenylated mRNAs are expressed in a developmentally regulated or tissue-specific manner [4, 24, 40, 41]. In addition, tissue-specific biases in alternative polyadenylation have been identified. For example, mRNAs with long 3′UTRs (selection of distal polyA sites) are most abundant in neural tissues, while mRNAs with short 3′UTRs (selection fo proximal polyA sites) are prevalent in testis. Important roles for alternative polyadenylation have been identified during T-cell stimulation [37], neuronal signaling [42], and in proliferation of tumor cell lines in culture [36]. In general, proliferating and undifferentiated cells tend to express mRNAs with short 3′UTRs while non-proliferating and/or differentiating cells (neurons and 'resting' T-cells, for example) generate mRNAs with long 3′UTRs [37, 43, 44]. Interestingly, 3′UTRs of a large number of germ cell mRNAs switch from long to short as cells progress through spermatogenesis [45–47], with

the selection of proximal polyA sites being a common feature of mRNAs expressed in round spermatids [48, 49]. It is not known whether accelerated decay of long 3′UTR mRNAs contributes to differences in the relative levels of long and short 3′UTR variants in different stages of spermatogenesis.

3.2 Functional Consequences of Alternative Processing of Germ Cell mRNAs

Compared to other tissues, the testis expresses higher numbers of alternatively spliced mRNAs including testis-specific mRNA variants, and mRNAs that exhibit stage-specific patterns of alternative splicing [23, 24, 50–52]. In addition to extensive stage-specific alternative splicing, changes in 3′UTRs caused by alternative polyadenylation are prevalent during spermatogenesis [45, 46].

The number of alternatively spliced mRNAs expressed in a given tissue generally correlate with the number of genes expressed (including those encoding splicing factors), suggesting that higher numbers of alternatively spliced mRNA variants result from increased combinations of splicing regulatory proteins [53]. In mice and humans, more genes are expressed in the testis (~84 % of RefSeq genes) than any other tissue [4]. Strikingly, the majority of RNA present in whole testis preparations is contributed by two germ cell types: pachytene spermatocytes (germ cells in meiotic prophase I where chromatin condensation and homologous recombination occurs) and round spermatids (the haploid products of meiosis). In these cells, a more open chromatin state facilitates promiscuous transcription of the genome including protein-coding and non-coding genes and intergenic elements (SINEs, LINEs, and LTRs) [54]. Collectively, these observations raise questions regarding the biologic importance of the expression of high numbers of alternatively spliced germ cell mRNAs and stage-specific changes in alternative splicing and polyadenylation during spermatogenesis. Nevertheless, specific examples of functional differences in alternatively processed germ cell mRNAs have been described.

3.2.1 LIG3, SOX17, and CREM

Representative examples of genes that yield alternative mRNAs that encode functionally distinct protein isoforms in mouse germ cells include LIG3, SOX17, and CREM. LIG3 encodes two isoforms (α and β) of DNA ligase III through utilization of distinct 3′ terminal exons. Both isoforms are highly expressed in testis, with DNA ligase III β mRNA being the predominant species. In somatic cells, both isoforms are expressed at low levels and DNA ligase III α mRNA is the predominant species. The α and β mRNAs yield polypeptides with different C-termini, and while both proteins are active as DNA joining enzymes, the β form (unlike the α form) is unable to interact with the DNA repair protein XRCC1, suggesting distinct cellular functions for the α and β isoforms of DNA ligase III [55, 56].

The SOX17 gene encodes a transcription factor bearing a high mobility group (HMG) box region in its N-terminus. In mouse testis, Sox17 is present in spermatogonia and has decreased expression in pachytene spermatocytes. Reduced Sox17 levels in pachytene cells is accompanied by increased expression of t-Sox17, an alternatively spliced variant that lacks the exon that codes for the majority of the HMG box region. As a result, t-Sox17 mRNA encodes a truncated protein which lacks an intact HMG box region and is unable to bind DNA or stimulate transcription of a luciferase reporter gene in co-transfection experiments [57].

Although the analysis of LIG3 and SOX17 mRNAs expressed in germ cells highlight the ability of alternative pre-mRNA processing to generate biochemically distinct polypeptides, the functional importance of these alternative isoforms (as with most germ cell alternative mRNAs that have been identified) has not been investigated *in vivo*. Transgenic mouse models that disrupt the balance of specific alternative mRNA variants or delay the expression of stage-specific isoforms are needed to determine whether specific changes in alternative mRNA isoforms are functionally important for mammalian germ cell development.

Studies of the transcription factor CREM (cAMP-responsive element modulator) highlight one of the best characterized examples of the importance of alternative pre-mRNA processing in spermatogenesis. Through selection of alternative promoters, alternative exon splicing, and alternative polyadenylation, the CREM gene can give rise to multiple mRNA and protein variants. In pre-meiotic germ cells and early prophase spermatocytes, the predominant CREM isoforms expressed (β and γ) are capable of binding CRE sequences of target genes but lack the glutamine-rich domains important for transactivation and thus function as suppressors of cAMP-induced transcription [58]. In pachytene spermatocytes, a switch in the pattern of CREM pre-mRNA splicing results in the production of CREMτ, which differs from β by the presence of two inserted glutamine-rich amino acid regions that confer transactivation function to CREMτ [59]. In addition to the conversion of CREM from suppressor to activator by alternative splicing, an alternative polyadenylation switch from distal to proximal polyA site use in the CREMτ 3′UTR eliminates multiple mRNA-destabilizaing elements and is associated with a robust increase in CREMτ mRNA levels in spermatids [60]. Thus, multiple changes in CREM pre-mRNA processing modulate protein function and mRNA abundance. Importantly, in mice homozygous for a null allele of CREM, spermatids fail to differentiate and there is an increase in germ cell apoptosis, suggesting that CREMτ is functionally important in transcriptional activation of genes in postmeiotic germ cells [61, 62].

3.2.2 Different Fates for Alternatively Polyadenylated Germ Cell mRNAs

In addition to CREMτ, several genes have been identified that yield alternative polyA variants in mouse testis with a bias towards selection of proximal sites (mRNAs with shorter 3′UTRs) in later stages of germ cell development [45, 62, 63]. The functional significance of 3′UTR shortening of large numbers of germ cell

mRNAs is not understood. Representative examples of genes that yield alternative 3′UTR variants with different cytoplasmic fates include RNF4 and DAZAP1. RNF4 encodes the small nuclear ring finger protein 4, a ubiquitin E3 ligase [64]. In spermatocytes and spermatids, two alternative RNF4 mRNA variants (of ~1.6 and 3.0 kb) are generated due to alternative use of different polyA sites present in the same 3′ terminal exon. Both the long and short 3′UTR mRNA isoforms of RNF4 are present at comparable levels in spermatocytes, while there is a significant increase in the abundance of the shorter ~1.6 kb RNF4 mRNA variant in spermatids [65]. Northern blot analysis of RNF4 mRNAs following sucrose density gradient centrifugation and fractionation of adult testis (a widely used method to assess the translational status of specific mRNAs) demonstrated that the long 3′UTR RNF4 isoform is polysome-associated while the short 3′UTR isoform is predominantly in the non-translating mRNP fraction. Thus, alternatively polyadenylated variants of RNF4 mRNA exhibit differences in ribosome association.

Differences in polysome-association for long and short 3′UTR variants have also been described for DAZAP1 mRNAs. DAZAP1 (DAZ associated protein 1) encodes a ubiquitously expressed RBP that is implicated in transcription, RNA splicing, and translation [66]. Two DAZAP1 mRNA isoforms generated by alternative polyadenylation at sites within the same 3′ terminal exon are present at comparable levels throughout testis development [67]. Both DAZAP1 mRNAs exhibit similar levels of polysome-association in early stages of postnatal testis development. However, as testis development proceeds in prepubertal mice and postmeiotic germ cells appear and gradually comprise a larger proportion of the total germ cell population, the short 3′UTR isoform is progressively reduced in the polysome fractions and localizes predominantly in the non-translating mRNP fractions. In contrast, the long 3′UTR DAZAP1 mRNA exhibits a modest decrease in polysome association with testis development, however the majority of the long 3′UTR isoform remains polysome-associated in adult testis. Thus, alternatively polyadenylated variants of DAZAP1 mRNA exhibit differences in ribosome-association in different stages of spermatogenesis.

Studies of alternative polyadenylation in T-cells and cancer cell lines in culture have posited that 3′UTR shortening due to selection of proximal alternative polyA sites functions as a mechanism to allow mRNAs to escape negative regulation imposed by elements present in long 3′UTRs [36]. In the examples presented here, the short 3′UTR variant of mRNAs derived from the RNF4 and DAZAP1 genes exhibited low levels of polysome-association in later stages of spermatogenesis. These observations indicate that 3′UTR shortening may not be obligatorily coupled to increased translation in all mammalian cell types. Multiple scenarios could account for the translational differences observed in these alternatively polyadenylated variants. One possibility is that reduced translation of the short 3′UTR isoforms in late stages of spermatogenesis results from the absence of specific ‘translation-promoting’ sequences that are present only in the long 3′UTR variants and may counteract ‘translation-repressing’ sequences present upstream of the proximal polyA site. The observation that the DAZAP1 short 3′UTR mRNA is competent for efficient translation in earlier stages of spermatogenesis, suggests the involvement

of stage-specific cofactors that determine if an mRNA is translationally active or repressed. Transgenic mouse models have shown that mRNAs expressed in spermatids can undergo sequence-independent assembly into translationally-repressed mRNP particles [68]. In addition, selection of proximal polyA sites is a common feature of mRNAs expressed in spermatids [45]. Thus, differences in the translation or repression of alternative polyA variants could be due, in part, to differences in the timing of their synthesis whereby widespread repression of the majority of mRNA is coincident with the synthesis of mRNAs with short 3′UTRs due to selection of proximal sites for 3′ end cleavage and polyadenylation. Insights into the molecular mechanisms and functional consequences of alternative polyadenylation during germ cell development awaits comprehensive measurements of the timing and dynamics of the synthesis of alternative mRNAs, their movement into and out of the translating and non-translating fractions, and their decay.

3.3 Translational Control: Global and Message-Specific

A third widespread form of post-transcriptional gene control in developing germ cells is the regulation of mRNA translation. This includes repression of the majority of germ cell mRNAs and translational activation of select mRNAs at specific stages of development. While the extent of repression varies between individual mRNA species, germ cell mRNAs generally show lower levels of polysome association compared to mRNAs expressed in somatic cells [69]. Nearly two-thirds of the total polyadenylated RNA present in isolated spermatocytes [70], spermatids [71], and in whole testis [72] is present in non-polysomal mRNP fractions, indicating that a significant portion of germ cell mRNAs are not involved in protein synthesis. Accordingly, a survey of eight tissues showed that the testis exhibits the lowest correlation between proteome and transcriptome, consistent with widespread translational repression of the majority of germ cell transcripts [73]. Global repression of mRNA translation is hypothesized to protect against over-production of proteins due to 'leaky or promiscuous' expression of large numbers of genes and higher mRNA levels compared to somatic cells [5, 54].

Regulation of mRNA translation has important roles throughout mammalian gametogenesis, from early steps of germ cell development during embryogenesis to release of spermatozoa into the lumen of seminiferous tubules in adults. For example, as germ cells transition from gonocytes to spermatogonia in the neonatal testis, ~50 genes show expression level changes while ~3000 mRNAs exhibit at least a twofold increase in translation efficiency [74, 75]. Thus, increased translation of a large cohort of mRNAs, rather than expansive changes in transcribed genes, remodels the germ cell proteome in the neonatal testis.

Stage-specific changes in mRNA translation have been described in multiple stages of spermatogenesis, but have been most intensively studied during spermatid differentiation [76, 77], where the assembly of mRNAs into translationally-repressed mRNPs and stage-specific release of specific mRNAs from this repression is essential. During

spermatid elongation, chromatin compaction results in the cessation of transcription, thus translation of new polypeptides necessary for the completion of spermatogenesis is dependent on mRNAs synthesized days earlier and stored in mRNPs [3, 78–80]. Examples of well-studied transcripts whose synthesis and translation are temporally disconnected include mRNAs encoding transition proteins (Tnp1 and Tnp2), protamine proteins (Prm1 and Prm2), and the sperm mitochondrial associated cysteine-rich protein (Smcp), all of which are encoded by genes that are essential for proper germ cell development in mice [81–85].

Prm1 and Tnp2 mRNAs are first detected in round spermatids but are not translated until several days later in elongating spermatids [69, 86–88]. Interestingly, these mRNAs only exhibit a partial release from translational repression (for example, less than half of the total Prm1 and Prm2 mRNAs become polysome-associated in elongating spermatids). The importance of UTR elements in temporal and stage-specific control of mRNA repression and translation was demonstrated in mice containing Prm1 or Tnp2 transgenes in which their respective UTRs were replaced, resulting in premature translation of Prm1 and Tnp2 mRNAs and germ cell developmental abnormalities [90, 91].

Transgenic mouse models have also provided key insights into potential regulatory mechanisms of mRNP assembly and stage-specific release of mRNAs from translational repression. To better understand mRNA assembly into and release from mRNPs in spermatids, a series of transgenic mice have been generated that contain a reporter gene (GFP or hGh) with various 5′ and/or 3′ UTRs, including those from translationally controlled mRNAs. Analyses of these transgenes revealed that mRNAs expressed in spermatids can undergo sequence-independent assembly into translationally-repressed mRNPs [68], and that specific sequences in the UTRs of Scmp and Prm1 participate in controlling the timing of mRNA release from mRNP particles and association with ribosomes [92–95]. Studies by Braun and colleagues identified two sequences in the Prm1 3′UTR that can delay translation of a reporter mRNA, including a translational control element (TCE) and a Y-box recognition sequence (YRS, UCCAUCA) that is recognized by Y-box proteins (discussed below) [94–97]. Interestingly, both the TCE and the YRS must be in close proximity to the polyA tail to function, however the molecular basis remains unknown.

Studies by Kleene and colleagues revealed a role for uORFs in translational control during spermiogenesis while demonstrating that both the 5′ and 3′ UTRs of Smcp are required to recapitulate the strength and duration of translational control observed with endogenous Smcp mRNAs [92, 93]. Additional evidence that 5′UTRs can influence translation of germ cell transcripts in spermatids comes from the analysis of alternative mRNAs derived from the AKAP4, TBP, and SOD1 genes. Selection of different sites of transcription initiation generates multiple alternative mRNA variants in germ cells from each of these genes. Interestingly, alternative transcription site use does not alter AKAP4, TBP, or SOD1 protein coding sequences, but does generate mRNAs that differ in their 5′UTRs and in translation efficiency [98–100]. These examples highlight an important role for alternative transcription start site selection in regulation of protein abundance.

3.4 Post-transcriptional Control Through PolyA Tail Length Regulation

The polyA tail present at the 3′end of nearly all mRNAs has important roles in several steps in the mRNA lifecycle including nuclear export, translational control, and mRNA stability [101, 102]. Many factors that impact post-transcriptional regulation of gene expression do so by directly or indirectly modifying polyA tail length. Deadenylation (polyA tail shortening) is the first step in the degradation of the majority of mRNAs [103]. Multiple post-transcriptional regulatory factors control their mRNA targets by affecting the recruitment and activity of deadenylases to specific mRNAs.

The importance of cytoplasmic polyA tail lengthening in mouse gametogenesis is illustrated by spermatogenic defects in knockout mice lacking a cytoplasmic polyA polymerase [104]. In addition, the cytoplasmic polyadenylation element binding protein (CPEB) important for cytoplasmic polyadenylation of specific mRNAs is essential in mouse germ cells. (discussed below, [105]). For some mRNAs, polyA tail lengthening in the cytoplasm occurs in response to specific cellular cues and is associated with increased mRNA translation. Translational activation resulting from polyA tail lengthening is believed to be due to stabilization of a circular 'closed loop' mRNA structure via interactions between polyA-binding protein (PABP) bound to the polyA tail and the eIF4E translation initiation complex bound to mRNA 5′ end 'cap' [101]. A requirement for 5′UTR and 3′UTR sequences in temporal control of translation in differentiating spermatids (discussed above) is consistent with a closed loop model of translational control of some germ cell mRNAs. Furthermore, mouse models have indicated that stage-specific modulation of PABP levels has an important role in the temporal control of translation and is critical for proper germ cell development [106].

A striking observation from northern blot analysis of translationally regulated mRNAs in spermatids is a difference in the electrophoretic mobility of specific mRNA species in ribosome-free mRNP and polysome-associated sucrose gradient fractions from mouse testis [86, 88, 107]. In the mRNP fractions, mRNAs encoding Prm1 or Tnp1 for example, appear as homogenous transcripts. However, the corresponding transcripts in the polysome fractions migrate as heterogeneous species or 'RNA smear' that results from polyA tail shortening. It remains unclear as to why translation of some mRNAs is accompanied by polyA tail shortening. Shortening does not appear to absolutely required however, as full length mRNAs (comparable in size to those in the mRNP fractions) do appear in the polysome fractions in addition to the isoforms with shorter polyA tails. It is not known whether partial deadenylation promotes release from translational repression, or if deadenylation occurs while mRNAs are being translated to act as a translational 'timer' whereby the polyA tail is progressively shortened until a critical length is reached that can no longer support interactions between the 5′ and 3′ ends and the mRNA.

4 Roles of RNA-binding Proteins in Germ Cell mRNA Regulation

The molecular mechanisms that control alternative mRNA expression and temporal control of translation during germ cell development remain poorly understood. Each step in the lifecycle of an mRNA is dependent on the combination and positions of bound RBPs. In the nucleus, co-transcriptional loading of specific RBPs onto nascent pre-mRNA can alter alternative splicing and alternative polyadenylation. In some cases, RBPs can function as either positive or negative-acting factors depending on their position relative to an alternative exon or polyA site [108]. RBPs also control downstream post-transcriptional events such as regulation of mRNA stability, localization, or translation. Thus, RBPs can regulate and integrate multiple layers of gene regulation to control which protein variants are made and modulate the timing, location, and dosage of mRNA translation. Accordingly, variations in the RBPs expressed in each cell type underlie tissue-specific differences in post-transcriptional gene regulation.

Additional modulation of post-transcriptional fate is achieved through changes in the in levels and/or activity of core RNA regulatory factors as well as auxiliary factors that do not directly bind RNA [108]. Core factors include those that are broadly expressed and have central roles in mRNA regulation. For example, members of the SR and hnRNP families of RBPs are widely expressed and generally function in activation or repression respectively, of exon splicing. Auxiliary factors modulate the activity of core factors through either direct binding to specific mRNAs (for example, miRNAs and tissue-specific RBPs) or indirectly through post-translational modifications of regulatory factors. An example of the latter includes members of the CLK/STY family of protein kinases that can phosphorylate SR proteins resulting in alterations in their localization and splicing activity [109–111]. Coincidentally, all four genes that encode CLK/STY kinases yield alternative mRNAs that are expressed in different tissue-specific combinations, with high levels of expression in the testis [112, 113]. It is not known whether differential expression of CLK/STY kinases contributes to stage specific changes in alternative splicing during spermatogenesis.

The conversion of an mRNA from a translationally-silent to active state is poorly understood, but likely to involve the selective removal and/or addition of specific RBPs [114, 115]. In male germ cells, remodeling of mRNPs is thought to have important roles in regulating the timing of mRNA stabilization, translation, and decay. RNA helicases are thought to reshape protein-RNA complexes—removing some RBPs and allowing others to bind [116]. Many RNA helicases localize to chromatoid bodies, dynamic spermatid-specific structures suggested to function as RNA processing centers involved in post-transcriptional RNA regulation [115, 117, 118].

Below, representative examples of RBPs believed to function as major contributors to post-transcriptional gene regulatory programs at specific stages of male gametogenesis are described. This includes factors that have either been shown to be essential for specific stages of germ cell development as well as RBPs that are suspected to have multiple key roles in post-transriptional control of germ cell mRNAs.

4.1 *Elavl1/HuR*

The embryonic lethal abnormal vision 1 protein (Elavl1, also known as HuR) is one of four related proteins of the Elavl/Hu family that have multiple roles in post-transcriptional regulation, including mRNA processing, export, translation, and decay [119]. Elavl1/HuR is broadly expressed, while the other members of the family (Elavl2/HuB, Elavl3/HuC, and Elavl3/HuD) are predominantly expressed in the nervous system. Elavl proteins bind their mRNA targets via interaction with U-rich sequences. In mouse brain and in mammalian cell culture, the Elavl RBPs binds thousands of transcripts via interactions in introns and 3′UTRs and regulate alternative mRNA splicing and mRNA stability [120–122]. In mouse spermatogenic cells, Elavl1/HuR is expressed in pachytene spermatocytes and round spermatids. Conditional inactivation of Elavl1 expression in germ cells results in male sterility due to defects in the completion of meiosis and failure of round spermatids to differentiate into elongated spermatids [123]. As a result, spermatozoa are absent in the epididymides. The direct role(s) of Elavl1 in mouse spermatogenic cells remain to be determined, however it is probable that Elavl1 is involved in multiple post-transcriptional regulatory events in germ cells since many regulatory roles have been attributed to Elavl proteins in different cellular contexts. To date, the functional consequence of Elavl1 deletion has been investigated on a single mRNA, HSP2A, that encodes a heat shock protein whose deletion results in a phenotype similar to Elavl1-null germ cells. Elavl1 binds HSP2A mRNA and loss of Elavl1 results in reduced levels of HSP2A mRNA on polysomes [123]. The molecular mechanism by which Elavl1 promotes HSP2A translation is not known.

4.2 *CELF (CUGBP, ELAV-Like Family) Proteins*

The CELF family of RBPs (CELF1-6; related to the ELAV family of RBPs) are multifunctional with roles in alternative splicing, mRNA translation, and mRNA deadenylation [124]. CELF proteins bind GU-rich elements (GREs). The roles of CELF proteins are dependent on the position of GRE-CELF interactions within mRNA targets, as has been demonstrated with other RBPs. Binding in the 5′UTR and 3′UTR has been linked to roles in mRNA translation and stability respectively, while binding within intronic sequences flanking alternative exons is associated with regulation of exon splicing [124, 125]. In mammals, CELF proteins have been shown to have important roles in regulation of developmentally regulated tissue-specific alternative mRNA splicing. In addition, CELF proteins have an evolutionarily conserved role in promoting mRNA decay through interactions with GREs in mRNA 3′UTRs [126]. In the mouse testis, CELF1 is expressed in both somatic cells and germ cells. In mixed background CELF1-null mice, a range of spermatogenic cell defects are observed including increased germ cell apoptosis and the absence of elongated spermatids [127]. While the direct roles of CELF proteins in

post-transcriptional control in germ cells are not known, the ability of CELF proteins to regulate multiple steps in mRNA metabolism suggest that CELF may also be multifunctional in mouse germ cells and coordinate pre-mRNA processing and cytoplasmic control of specific mRNAs.

4.3 Sam68

Sam68 is a member of the STAR (signal transduction and activation of RNA) family of RBPs that link signal transduction pathways to post-transcriptional regulation of mRNA [128]. In response to activation of signaling pathways, Sam68 can be phosphorylated resulting in a change in Sam68 subcellular localization and/or its activity on its mRNA targets. For example, depolarization of neurons results in the activation of the calcium/calmodulin dependent kinase IV (which has an important role in activity-dependent alternative mRNA splicing of many neuronal pre-mRNAs) as a result of the phosphorylation of Sam68 [129]. Among the mRNAs whose splicing is altered in an activity-dependent manner include those that encode the synaptic receptors Neurexin-1, -2, and -3 (encoded by three separate genes yet capable of yielding ~1000 mRNA isoforms due to extensive alternative splicing) [130]. Sam68 regulates Neurexin pre-mRNA splicing *in vitro*, while the absence of Sam68 in vivo results in the failure of neurexin pre-mRNAs to be spliced in response to neuronal activation [129]. Collectively, these observations indicate that Sam68 is an important mediator of activity-dependent changes in mRNA processing in the brain.

In the testis, Sam68 is expressed in spermatogenic cells and Sertoli cells [131]. Interestingly, Sam68 localization within germ cells differs in different cell types with nuclear expression in spermatogonia, pachytene spermatocytes and round spermatids, and cytoplasmic localization during meiotic divisions where it associates with polysomes [132]. Translocation of Sam68 to the cytoplasm coincides with its phosphorylation. Sam68 knockout mice are infertile and exhibit a range of spermatogenic defects including high numbers of apoptotic cells, aberrant nuclear divisions, and misshapen and immotile spermatozoa [133]. Multiple defects in mRNA regulation have been identified in Sam68 null testis. This includes up-regulation and down-regulation of ~100 and ~300 genes respectively, decreased polysome-association of specific mRNAs [133], and aberrant alternative mRNA splicing [134]. Thus, Sam68 is a multifunctional RBP in male germ cell development. Interestingly, a second member of the STAR family of RBPs, T-STAR (also called SLM2), is highly expressed in testis yet is dispensable for spermatogenesis. The overlapping expression of Sam68 and T-STAR in mouse germ cells suggests that Sam68 may functionally compensate for the loss of T-STAR, however T-STAR is not able to compensate for the loss of Sam68 in spermatogenic cells [131].

4.4 PTB (Polypyrimidine Tract Binding) Family of RBPs

The PTB family of RBPs includes Ptbp1 (more commonly known as PTB), Ptbp2 (also called brain or neuronal PTB, brPTB and nPTB respectively), and Ptbp3 (also called ROD1). Studies in several model systems (from *in vitro* assays to mammalian tissue) have identified multiple roles for Ptbp1 in mRNA regulation, including control of mRNA splicing, mRNA stability, and localization (for review, see [135]). While the role of PTB proteins in mammalian germ cell development have not been described, multiple lines of evidence suggest that they are likely to have important roles in post-transcriptional control of germ cell mRNAs.

In the mouse testis, Ptbp1 expression is restricted to spermatogonia while Ptbp2 is expressed in spermatocytes and spermatids [136]. The reciprocal expression of Ptbp1 and Ptbp2 in different phases of spermatogenesis suggests that Ptbp1 and Ptbp2 have distinct roles in different stages of germ cell development. Recent analyses of sequences associated with alternative exons that are differentially spliced in 6- and 21-day old testis (where the most advanced germ cells are spermatogonia and spermatids, respectively) revealed an enrichment of motifs that match binding sites for PTB proteins, suggesting that one or both of the PTB proteins expressed in different stages of spermatogenesis may have important roles in temporal control of germ cell mRNA splicing [137]. In the embryonic brain Ptbp2 functions predominantly as a silencer of alternative exon splicing through interactions upstream of and/or within alternative exons [138]. Furthermore, exons that are repressed by Ptbp2 in the embryonic brain correspond to exons that are activated (spliced) in adult brain. Whether Ptbp2 has a similar role in temporal control of alternative exons during germ cell development remains to be determined.

A splicing-independent role for PTB proteins in germ cell post-transcriptional control has also been proposed. In one report, incubation of *in vitro* synthesized RNA probes in mouse testis lysate identified Ptbp2 binding to a specific region of the 3′UTR of Pgk2 mRNA, an mRNA that is first detected in meiotic cells whereas PGK2 protein is first detected in post-meiotic cells [136]. In HeLa cells and *in vitro*, Ptbp2 was able to increase PGK2 mRNA half-life suggesting a role for Ptbp2 in stabilization of this mRNA during spermatogenesis. mRNA stabilization is an important component of temporal mRNA control, ensuring that mRNAs are available as templates for translation in mid- to late-spermiogenesis [78–80]. Whether Ptbp2 regulates the stability of Pgk2 and other mRNAs in germ cells is not known and will require comparative analysis of mRNA steady state levels in wild type and Ptbp2-null testes.

4.5 τ-Cstf64

τ-Cstf64 (encoded by CSTF2T) is a retrotransposed paralog of the CSTF2 gene that encodes the 64 kilodalton subunit of the CSTF (cleavage stimulatory factor) complex required for 3′ end cleavage and polyadenylation. Due to the location of CSTF2

on the X chromosome, meiotic sex chromosome inactivation results in a loss of Cstf64 expression in pachytene spermatocytes. τ-Cstf64 is expressed in pachytene spermatocytes (where it is proposed to functionally compensate for the loss of Cstf64) and continues through the early spermatid stages [139]. Interestingly, expression of τ-Cstf64 coincides with increased selection of proximal alternative polyadenylation sites, which generally contain a non-canonical polyA signal [43, 45]. Together, these observations led to the hypothesis that selection of proximal polyA sites in germ cells results from differences in the relative levels and activity of Cst64 and τ-Cstf64. However, recent analyses indicate that Cst64 and τ-Cstf64 have highly similar RNA-binding specificities and overlapping functionality [140–142], indicating that other factors and regulatory mechanisms may contribute to changes in the alternative polyA site selection during spermatogenesis. For example, inactivation of the gene encoding BRDT, a member of the BET (bromodomain and extra terminal motif) family of chromatin-interacting regulators of transcription, results in reduced accumulation of mRNAs processed at proximal polyA sites, indicating that alternative polyA site selection in developing germ cells may be coupled to transcriptional activity and/or chromatin state [143]. Nonetheless, τ-Cstf64 expression is necessary for proper germ cell development as several spermatogenic defects are observed in CSTF2T-null mice [144, 145].

4.6 Y-Box Proteins

The Y-box proteins are functionally conserved DNA and RNA-binding proteins with important roles in binding and regulating mRNAs in germ cells [146, 147, 148]. Three separate Y-box genes are present in mice: YBX1 (also called MSY1), YBX2 (MSY2), and YBX3 (MSY4) which yields two protein isoforms (long and short, YBX3L and YBX3S respectively) that are expressed at comparable levels in mouse testis and are derived from alternative exon splicing. In mouse testis, Y-box proteins are found in association with translationally repressed mRNP fractions [146, 149]. Approximately 75 % of testis polyadenylated RNA is complexed with YBX2 and YBX3. YBX2 and YBX3 are expressed in meiotic and post-meiotic cells and are essential for spermatogenesis [146–148]. Interestingly, transgenic mice engineered to prolong YBX3 expression beyond the round spermatid stage interfered with translational activation of temporal controlled mRNAs bearing a 3′UTR Y-box recognition sequence [146]. Collectively, these studies indicate that Y-box proteins mediate the storage and masking of mRNAs during mouse spermatogenesis.

4.7 CPEB (Cytoplasmic Polyadenylation Element Binding Protein)

The cytoplasmic polyadenylation element (CPE) is a U-rich 3′UTR sequence (usually $U_{4–5}A_{1–2}$,U) that has important roles in translational regulation. The CPE is recognized by CPE-binding protein (CPEB), which has dual roles in translation by

acting as a central component of regulatory pathways that promote or repress mRNA translation. In response to phosphorylation, CPEB bound to a CPE that is close to a polyA signal element (usually AAUAAA located 10–30 bases upstream of the site of cleavage) can promote the formation of an active cytoplasmic polyadenylation complex, resulting in polyA tail lengthening and activation of translation [150]. CPEB bound to a CPE can repress translation through interactions with factors that prevent translation initiation or recruit components of the deadenylation and mRNA decay machinery [101, 151]. The specific action of CPEB on a given mRNA (as activator or repressor) as well as the magnitude of translational regulation is dictated by the combination and positions of CPEs as well as binding elements for other RBPs in the 3′UTR. For example, the presence of a PBE (UGUANAUA, the binding site for the Pumilio family of RBPs, discussed below) can enhance CPEB-mediated translational activation with additional PBEs exerting a positive effect [152]. Combinatorial regulation of CPE-containing mRNAs allows for coordinated control of networks of co-regulated mRNAs and ensures that not all CPE-containing mRNAs are repressed and/or translated at the same time. Additional regulatory complexity can be achieved by recruitment of CPEB to mRNAs that lack detectable CPEs via interactions with other RBPs bound to their cognate binding sites in a 3′UTR [153].

The importance of CPEB in mouse germ cells is demonstrated by spermatogenic arrest in CPEB1-null mice. In the absence of CPEB1 (one of four mouse CPEB genes), male germ cells contain fragmented and dispersed chromatin, consistent with defects in chromosome synapsis and recombination [105]. Importantly, the SYCP1 and SYCP3 mRNAs encoding components of the synaptomenal complex required for pairing of sister chromatids, contain CPEs in their 3′UTRs. In mouse oocytes, SYCP1 and SYCP3 mRNA levels are unaffected by the loss of CPEB1, however their polyA tail lengths are reduced and the corresponding proteins absent [105]. These observations suggest a direct role for CPEB1 in the regulation of polyA tail length and production of synaptonemal complex proteins in mouse spermatogenesis. Interestingly, mouse CPEB paralogs display overlapping yet distinct patterns of expression during spermatogenesis [105, 154] raising the possibility that additional complexity of CPEB-mediated control via specific CPEB isoforms.

4.8 Pumilio and Nanos

The Pumilio (PUM) proteins are members of the PUF (Pumilio and FBF) family of RBPs that are structurally and functionally conserved from yeast to)mammals and plants [155, 156]. PUM proteins)function as critical post-transcriptional regulators of cell fate and developmental programs of gene expression. PUM proteins regulate their mRNA targets through interactions with one or more PUM binding elements (PBEs, with the consensus sequence UGUANAUA) present in the 3′UTR. While positive roles for PUM proteins have been described in CPEB-mediated translation regulation (see above), most analyses of PUM proteins have focused on their roles in translational repression and de-stabilization of specific mRNAs.

The mouse and human genomes each contain genes for two PUM proteins (Pum1 and Pum2) with widespread and overlapping expression in different tissues [155]. Pum2 appears to be dispensable for male germ cell development in mice, as fertility is unaffected in animals that are homozygous for a genetrap mutation that abrogates Pum2 expression [157]. In contrast, Pum1 knockout mice exhibit spermatogenic defects including reduced numbers of spermatozoa and increased numbers of apoptotic spermatocytes [158]. Analysis of RNAs that co-precipitate with Pum1 from adult testis provided insights into the molecular basis of the Pum1-null testis phenotype. Among the 1527 genes whose mRNA products are bound by Pum1, is an enrichment of genes encoding regulators of the cell cycle and p53. In the absence of Pum1, mRNAs encoding eight activators of p53 are upregulated resulting in activation of p53. Thus, Pum1 appears to have a role in safeguarding spermatogenic cells from apoptotic programs.

mRNA regulation by PUM proteins involves direct and indirect interactions with components of mRNA degradation complexes. For example, physical interactions between PUF proteins and components of the deadenylation machinery have been described from yeast to human [158]. Deadenylation factors can also be indirectly recruited to mRNAs with PUM bound to a 3′UTR PBE through other PUM-interacting factors. For example, the Nanos proteins are evolutionarily conserved post-transcriptional regulatory factors that interact with and are dependent on PUM in order to be recruited to specific mRNAs. Biochemical purification of Nanos proteins from mouse testis identified several components of the CCR4-NOT deadenylation complex as co-purified proteins, indicating that Nanos proteins down-regulate their mRNA targets through recruitment of the mRNA deadenylation complex [159].

Gene knockouts and transgenic strains have demonstrated essential roles for two mouse Nanos genes (Nanos2 and Nanos3) in mouse germ cell development [160]. In the absence of Nanos3, PGCs that have migrated to the genital ridge exhibit progressive reductions in number and eventually are all lost [161]. A similar failure of PGCs to proliferate and/or survive has also been described in mouse knockouts of the RBPs TIAL1 and DND1 [162, 163]. Germ cell loss also occurs in Nano2-deficient mice [161]. In spermatogonia, Nanos2 has a critical role in regulating the balance between self-renewal and differentiation, as postnatal deletion of Nanos2 results in depletion of undifferentiated spermatogonia, whereas Nanos2 overexpression results in the accumulation of undifferentiated spermatogonia [164].

4.9 Dazl

Dazl (Deleted in Azoospermia [Daz] -Like] is the autosomal homolog of the primate-specific Daz gene (deleted in azoospermia) present on chromosome)Y and deleted in a subset of men with spermatogenic failure (~5–10 %; ranging from complete)absence of germ cells to oligozoospermia) [165, 166]. Daz, Dazl, and Boule comprise the Daz family of proteins, all of which have a single RRM-type RNA-binding protein, have overlapping yet distinct expression patterns in male germ

cells, and are required for germ cell development in a variety of model organisms [166]. Importantly, the human Daz gene can partially rescue defects in Dazl-null mice, indicating that these highly conserved proteins have considerable functional conservation [167]. In mice with a mixed genetic background (CD1 and C57BL/6J), Dazl-null testes exhibit reduced numbers of germ cells, most of which appear to arrest at early stages of spermatogonial proliferation and differentiation [168–170]. The few Dazl-null germ cells that do enter meiosis, arrest in an early (pre-leptotene) stage of meiosis I. In a pure C57BL/6J genetic background, Dazl-null germ cells exhibit earlier defects including impaired primordial germ cell development and migration [171].

Multiple lines of evidence (reviewed in [166]) support a role for Dazl as a positive regulator of mRNA translation via interactions with Dazl RNA-binding sites (short polyU-rich tracts interspersed by G nucleotides, or C to a lesser extent) in the 3′UTR of target mRNAs. Of the ~20 genes whose mRNA products either co-purify with or interact with Dazl in mouse testis lysate, few have been directly assayed for protein levels in Dazl-null germ cells. For two target genes (Sycp3 and Mvh) immunofluorescence assays demonstrate reduced protein levels, although the magnitude of the reduction is variable between cells [172, 173]. Positive roles for Dazl in translation have also been observed using reporter mRNAs in oocytes and transfected cells [166]. Interestingly, in some cases identical reporter mRNA (containing a 3′UTR from a Dazl mRNA target) exhibits opposite responses to co-transfection of Dazl-expressing constructs in different cell lines. For example, lacZ reporter constructes bearing the Mvh 3′UTR show increased translation when co-transfected with Dazl expression vector in GC-1 cells, and reduced levels compared to no Dazl control cells in embryonic stem cells [174]. These findings indicate that Dazl may also have a negative role in translation and that cell context may be a determinant of Dazl function. Additional evidence of multiple roles for Dazl in post-transcriptional regulation is suggested by the presence of Dazl in both non-translating and polysomal fractions prepared from adult testis [175], the accumulation of Dazl in stress granules following heat stress of mouse testis, and a requirement of Dazl for such granules to form [176].

5 Conclusion

The importance of post-transcriptional control in mammalian germ cell development is evident from the large number of mouse models that have identified essential roles for specific RBPs at every stage of germ cell development, from the survival and proliferation of embryonic germ cells to the release of mature spermatozoa into the lumen of seminiferous tubules. In addition, transgenic mice have demonstrated the necessity of stage-specific post-transcriptional control, and revealed specific sequences in mRNAs that are essential for such regulation. Despite significant progress in identifying RBPs and post-transcriptional regulatory events that are essential for germ cell development, significant questions and

challenges remain. What mRNAs are directly bound by specific RBPs and what are the functional consequences of such interactions? How are networks of co-regulated mRNAs controlled at specific stages of development? How are changes in alternative mRNA isoform expression regulated, and what are the functional consequences of such regulation? While cell culture and *in vitro* assays can provide some information, the inability of such approaches to recapitulate regulatory events or accurately predict roles of specific RBPs *in vivo* (as described by Kleene [6]) limits their utility. Continued use of transgenic and knockout models in combination with cellular and biochemical enrichment tools and high throughput sequencing methodologies (including CLIP, RNA-Seq, and ribosome profiling, for example) promises to provide new insights into mechanisms of stage-specific post-transcriptional regulation in a transcriptome-wide manner. The demonstration of essential roles for germ cell RBPs in an array of cellular processes including control of self renewal, proliferation, entry into meiosis and differentiation highlights mouse spermatogenesis as a powerful model system to investigate how post-transcriptional controls drive mammalian cell development *in vivo*.

References

1. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793
2. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463
3. Kleene KC (2003) Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. Cytogenet Genome Res 103:217–224
4. Ramskold D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol 5, e1000598
5. Kleene KC (2001) A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. Mech Dev 106:3–23
6. Kleene KC (2013) Connecting cis-elements and trans-factors with mechanisms of developmental regulation of mRNA translation in meiotic and haploid mammalian spermatogenic cells. Reproduction 146:R1–R19
7. Cooke HJ, Saunders PT (2002) Mouse models of male infertility. Nat Rev Genet 3:790–801
8. Jamsai D, O'Bryan MK (2011) Mouse models in male fertility research. Asian J Androl 13:139–151
9. Rossi P, Dolci S (2013) Paracrine Mechanisms Involved in the Control of Early Stages of Mammalian Spermatogenesis. Front Endocrinol (Lausanne) 4:181
10. Ginsburg M, Snow MH, McLaren A (1990) Primordial germ cells in the mouse embryo during gastrulation. Development 110:521–528
11. McLaren A (2001) Mammalian germ cells: birth, sex, and immortality. Cell Struct Funct 26:119–122
12. Nagano R, Tabata S, Nakanishi Y, Ohsako S, Kurohmaru M, Hayashi Y (2000) Reproliferation and relocation of mouse male germ cells (gonocytes) during prespermatogenesis. Anat Rec 258:210–220
13. Kopera IA, Bilinska B, Cheng CY, Mruk DD (2010) Sertoli-germ cell junctions in the testis: a review of recent data. Philos Trans R Soc Lond B Biol Sci 365:1593–1605
14. de Rooij DG, Griswold MD (2012) Questions about spermatogonia posed and answered since 2000. J Androl 33:1085–1095
15. Song HW, Wilkinson MF (2014) Transcriptional control of spermatogonial maintenance and differentiation. Semin Cell Dev Biol 30:14–26

16. Yoshida S (2008) Spermatogenic stem cell system in the mouse testis. Cold Spring Harb Symp Quant Biol 73:25–32
17. Youds JL, Boulton SJ (2011) The choice in meiosis—defining the factors that influence crossover or non-crossover formation. J Cell Sci 124:501–513
18. Gupta SK, Bhandari B (2011) Acrosome reaction: relevance of zona pellucida glycoproteins. Asian J Androl 13:97–105
19. Monesi V, Geremia R, D'Agostino A, Boitani C (1978) Biochemistry of male germ cell differentiation in mammals: RNA synthesis in meiotic and postmeiotic cells. Curr Top Dev Biol 12:11–36
20. Kierszenbaum AL, Tres LL (1975) Structural and transcriptional features of the mouse spermatid genome. J Cell Biol 65:258–270
21. Upadhyay RD, Kumar AV, Ganeshan M, Balasinor NH (2012) Tubulobulbar complex: cytoskeletal remodeling to release spermatozoa. Reprod Biol Endocrinol 10:27
22. Gilbert W (1978) Why genes in pieces? Nature 271:501
23. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415
24. Wang ET, Sandberg R, Luo S et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476
25. Barbosa-Morais NL, Irimia M, Pan Q et al (2012) The evolutionary landscape of alternative splicing in vertebrate species. Science 338:1587–1593
26. Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338:1593–1599
27. Miura K, Fujibuchi W, Unno M (2012) Splice variants in apoptotic pathway. Exp Oncol 34:212–217
28. Zhang C, Frias MA, Mele A et al (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science 329:439–443
29. Ellis JD, Barrios-Rodiles M, Colak R et al (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell 46:884–892
30. Buljan M, Chalancon G, Eustermann S et al (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. Mol Cell 46:871–883
31. McGlincy NJ, Smith CW (2008) Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci 33:385–393
32. Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. Genes Dev 25:1770–1782
33. Bentley DL (2014) Coupling mRNA processing with transcription in time and space. Nat Rev Genet 15:163–175
34. Shi Y (2012) Alternative polyadenylation: new insights from global analyses. RNA 18:2105–2117
35. Rouget C, Papin C, Boureux A et al (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467:1128–1132
36. Mayr C, Bartel DP (2009) Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138:673–684
37. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science 320:1643–1647
38. Ghosh T, Soni K, Scaria V, Halimani M, Bhattacharjee C, Pillai B (2008) MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic {beta}-actin gene. Nucleic Acids Res 36:6318–6332
39. Mignone F, Gissi C, Liuni S, Pesole G (2002) Untranslated regions of mRNAs. Genome Biol 3, REVIEWS0004
40. Tian B, Hu J, Zhang H, Lutz CS (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33:201–212
41. Tian B, Manley JL (2013) Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci 38:312–320

42. Flavell SW, Kim TK, Gray JM et al (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60:1022–1038
43. Zhang H, Lee JY, Tian B (2005) Biased alternative polyadenylation in human tissues. Genome Biol 6:R100
44. Ji Z, Lee JY, Pan Z, Jiang B, Tian B (2009) Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A 106:7028–7033
45. Liu D, Brockman JM, Dass B et al (2007) Systematic variation in mRNA 3′-processing signals during mouse spermatogenesis. Nucleic Acids Res 35:234–246
46. McMahon KW, Hirsch BA, MacDonald CC (2006) Differences in polyadenylation site choice between somatic and male germ cells. BMC Mol Biol 7:35
47. Wang H, Sartini BL, Millette CF, Kilpatrick DL (2006) A developmental switch in transcription factor isoforms during spermatogenesis controlled by alternative messenger RNA 3′-end formation. Biol Reprod 75:318–323
48. O'Brien DA, Welch JE, Fulcher KD, Eddy EM (1994) Expression of mannose 6-phosphate receptor messenger ribonucleic acids in mouse spermatogenic and Sertoli cells. Biol Reprod 50:429–435
49. Shaper NL, Wright WW, Shaper JH (1990) Murine beta 1,4-galactosyltransferase: both the amounts and structure of the mRNA are regulated during spermatogenesis. Proc Natl Acad Sci U S A 87:791–795
50. Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. Genome Biol 5:R74
51. Clark TA, Schweitzer AC, Chen TX et al (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol 8:R64
52. Grosso AR, Gomes AQ, Barbosa-Morais NL et al (2008) Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res 36:4823–4832
53. de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D (2010) Splicing factor and exon profiling across human tissues. Nucleic Acids Res 38:2825–2838
54. Soumillon M, Necsulea A, Weier M et al (2013) Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3:2179–2190
55. Mackey ZB, Ramos W, Levin DS, Walter CA, McCarrey JR, Tomkinson AE (1997) An alternative splicing event which occurs in mouse pachytene spermatocytes generates a form of DNA ligase III with distinct biochemical properties that may function in meiotic recombination. Mol Cell Biol 17:989–998
56. Nash RA, Caldecott KW, Barnes DE, Lindahl T (1997) XRCC1 protein interacts with one of two distinct forms of DNA ligase III. Biochemistry 36:5207–5211
57. Kanai Y, Kanai-Azuma M, Noce T et al (1996) Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis. J Cell Biol 133:667–681
58. Foulkes NS, Borrelli E, Sassone-Corsi P (1991) CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. Cell 64:739–749
59. Foulkes NS, Mellstrom B, Benusiglio E, Sassone-Corsi P (1992) Developmental switch of CREM function during spermatogenesis: from antagonist to activator. Nature 355:80–84
60. Foulkes NS, Schlotter F, Pevet P, Sassone-Corsi P (1993) Pituitary hormone FSH directs the CREM functional switch during spermatogenesis. Nature 362:264–267
61. Nantel F, Monaco L, Foulkes NS et al (1996) Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. Nature 380:159–162
62. Blendy JA, Kaestner KH, Weinbauer GF, Nieschlag E, Schutz G (1996) Severe impairment of spermatogenesis in mice lacking the CREM gene. Nature 380:162–165
63. Yang G, Zhang YL, Buchold GM, Jetten AM, O'Brien DA (2003) Analysis of germ cell nuclear factor transcripts and protein expression during spermatogenesis. Biol Reprod 68:1620–1630
64. Groocock LM, Nie M, Prudden J et al (2014) RNF4 interacts with both SUMO and nucleosomes to promote the DNA damage response. EMBO Rep 15:601–608

65. Pero R, Lembo F, Chieffi P et al (2003) Translational regulation of a novel testis-specific RNF4 transcript. Mol Reprod Dev 66:1–7
66. Hsu LC, Chen HY, Lin YW et al (2008) DAZAP1, an hnRNP protein, is required for normal growth and spermatogenesis in mice. RNA 14:1814–1822
67. Yang CK, Yen P (2013) Differential translation of Dazap1 transcripts during spermatogenesis. PLoS One 8, e60873
68. Schmidt EE, Hanson ES, Capecchi MR (1999) Sequence-independent assembly of spermatid mRNAs into messenger ribonucleoprotein particles. Mol Cell Biol 19:3904–3915
69. Cataldo L, Mastrangelo MA, Kleene KC (1999) A quantitative sucrose gradient analysis of the translational activity of 18 mRNA species in testes from adult mice. Mol Hum Reprod 5:206–213
70. Gold B, Stern L, Bradley FM, Hecht NB (1983) Gene expression during mammalian spermatogenesis. II. Evidence for stage-specific differences in mRNA populations. J Exp Zool 225:123–134
71. Stern L, Kleene KC, Gold B, Hecht NB (1983) Gene expression during mammalian spermatogenesis. III. Changes in populations of mRNA during spermiogenesis. Exp Cell Res 143:247–255
72. Gold B, Hecht NB (1981) Differential compartmentalization of messenger ribonucleic acid in murine testis. Biochemistry 20:4871–4877
73. Cagney G, Park S, Chung C et al (2005) Human tissue profiling with multidimensional protein identification technology. J Proteome Res 4:1757–1767
74. Shima JE, McLean DJ, McCarrey JR, Griswold MD (2004) The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. Biol Reprod 71:319–330
75. Chappell VA, Busada JT, Keiper BD, Geyer CB (2013) Translational activation of developmental messenger RNAs during neonatal mouse testis development. Biol Reprod 89:61
76. Iguchi N, Tobias JW, Hecht NB (2006) Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. Proc Natl Acad Sci U S A 103:7712–7717
77. Braun RE (1998) Post-transcriptional control of gene expression during spermatogenesis. Semin Cell Dev Biol 9:483–489
78. Steger K (2001) Haploid spermatids exhibit translationally repressed mRNAs. Anat Embryol (Berl) 203:323–334
79. Braun RE (2000) Temporal control of protein synthesis during spermatogenesis. Int J Androl 23(Suppl 2):92–94
80. Hecht NB (1998) Molecular mechanisms of male germ cell differentiation. Bioessays 20:555–561
81. Yu YE, Zhang Y, Unni E et al (2000) Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc Natl Acad Sci U S A 97:4683–4688
82. Zhao M, Shirley CR, Yu YE et al (2001) Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. Mol Cell Biol 21:7243–7255
83. Adham IM, Nayernia K, Burkhardt-Gottges E et al (2001) Teratozoospermia in mice lacking the transition protein 2 (Tnp2). Mol Hum Reprod 7:513–520
84. Cho C, Willis WD, Goulding EH et al (2001) Haploinsufficiency of protamine-1 or -2 causes infertility in mice. Nat Genet 28:82–86
85. Nayernia K, Adham IM, Burkhardt-Gottges E et al (2002) Asthenozoospermia in mice with targeted deletion of the sperm mitochondrion-associated cysteine-rich protein (Smcp) gene. Mol Cell Biol 22:3046–3052
86. Kleene KC (1989) Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. Development 106:367–373
87. Mali P, Kaipia A, Kangasniemi M et al (1989) Stage-specific expression of nucleoprotein mRNAs during rat and mouse spermiogenesis. Reprod Fertil Dev 1:369–382
88. Kleene KC, Distel RJ, Hecht NB (1984) Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. Dev Biol 105:71–79
89. Shih DM, Kleene KC (1992) A study by in situ hybridization of the stage of appearance and disappearance of the transition protein 2 and the mitochondrial capsule seleno-protein mRNAs during spermatogenesis in the mouse. Mol Reprod Dev 33:222–227

90. Lee K, Haugen HS, Clegg CH, Braun RE (1995) Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. Proc Natl Acad Sci U S A 92:12451–12455
91. Tseden K, Topaloglu O, Meinhardt A et al (2007) Premature translation of transition protein 2 mRNA causes sperm abnormalities and male infertility. Mol Reprod Dev 74:273–279
92. Bagarova J, Chowdhury TA, Kimura M, Kleene KC (2010) Identification of elements in the Smcp 5′ and 3′ UTR that repress translation and promote the formation of heavy inactive mRNPs in spermatids by analysis of mutations in transgenic mice. Reproduction 140:853–864
93. Hawthorne SK, Busanelli RR, Kleene KC (2006) The 5′ UTR and 3′ UTR of the sperm mitochondria-associated cysteine-rich protein mRNA regulate translation in spermatids by multiple mechanisms in transgenic mice. Dev Biol 297:118–126
94. Fajardo MA, Haugen HS, Clegg CH, Braun RE (1997) Separate elements in the 3′ untranslated region of the mouse protamine 1 mRNA regulate translational repression and activation during murine spermatogenesis. Dev Biol 191:42–52
95. Zhong J, Peters AH, Kafer K, Braun RE (2001) A highly conserved sequence essential for translational repression of the protamine 1 messenger rna in murine spermatids. Biol Reprod 64:1784–1789
96. Giorgini F, Davies HG, Braun RE (2001) MSY2 and MSY4 bind a conserved sequence in the 3′ untranslated region of protamine 1 mRNA in vitro and in vivo. Mol Cell Biol 21:7010–7019
97. Braun RE, Peschon JJ, Behringer RR, Brinster RL, Palmiter RD (1989) Protamine 3′-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. Genes Dev 3:793–802
98. Nipper RW, Chennothukuzhi V, Tutuncu L, Williams CJ, Gerton GL, Moss SB (2005) Differential RNA expression and polyribosome loading of alternative transcripts of the Akap4 gene in murine spermatids. Mol Reprod Dev 70:397–405
99. Schmidt EE, Schibler U (1997) Developmental testis-specific regulation of mRNA levels and mRNA translational efficiencies for TATA-binding protein mRNA isoforms. Dev Biol 184:138–149
100. Gu W, Morales C, Hecht NB (1995) In male mouse germ cells, copper-zinc superoxide dismutase utilizes alternative promoters that produce multiple transcripts with different translation potential. J Biol Chem 270:236–243
101. Weill L, Belloc E, Bava FA, Mendez R (2012) Translational control by changes in poly(A) tail length: recycling mRNAs. Nat Struct Mol Biol 19:577–585
102. Eckmann CR, Rammelt C, Wahle E (2011) Control of poly(A) tail length. Wiley Interdiscip Rev RNA 2:348–361
103. Doidge R, Mittal S, Aslam A, Winkler GS (2012) Deadenylation of cytoplasmic mRNA by the mammalian Ccr4-Not complex. Biochem Soc Trans 40:896–901
104. Kashiwabara S, Noguchi J, Zhuang T et al (2002) Regulation of spermatogenesis by testis-specific, cytoplasmic poly(A) polymerase TPAP. Science 298:1999–2002
105. Tay J, Richter JD (2001) Germ cell differentiation and synaptonemal complex formation are disrupted in CPEB knockout mice. Dev Cell 1:201–213
106. Yanagiya A, Delbes G, Svitkin YV, Robaire B, Sonenberg N (2010) The poly(A)-binding protein partner Paip2a controls translation during late spermiogenesis in mice. J Clin Invest 120:3389–3400
107. Gu W, Kwon YK, Hecht NB (1996) In postmeiotic male germ cells poly (A) shortening accompanies translation of mRNA encoding gamma enteric actin but not cytoplasmic beta and gamma actin mRNAs. Mol Reprod Dev 44:141–145
108. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11:75–87
109. Colwill K, Pawson T, Andrews B et al (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J 15:265–275
110. Duncan PI, Stojdl DF, Marius RM, Bell JC (1997) In Vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. Mol Cell Biol 17:5996–6001

111. Prasad J, Colwill K, Pawson T, Manley JL (1999) The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. Mol Cell Biol 19:6991–7000
112. Howell BW, Afar DE, Lew J et al (1991) STY, a tyrosine-phosphorylating enzyme with sequence homology to serine/threonine kinases. Mol Cell Biol 11:568–572
113. Nayler O, Stamm S, Ullrich A (1997) Characterization and comparison of four serine- and arginine-rich (SR) protein kinases. Biochem J 326:693–700
114. Arkov AL, Ramos A (2010) Building RNA-protein granules: insight from the germline. Trends Cell Biol 20:482–490
115. Kotaja N, Sassone-Corsi P (2007) The chromatoid body: a germ-cell-specific RNA-processing centre. Nat Rev Mol Cell Biol 8:85–90
116. Putnam AA, Jankowsky E (1829) DEAD-box helicases as integrators of RNA, nucleotide and protein binding. Biochim Biophys Acta 2013:884–893
117. Kleene KC, Cullinane DL (2011) Maybe repressed mRNAs are not stored in the chromatoid body in mammalian spermatids. Reproduction 142:383–388
118. Meikar O, Da Ros M, Korhonen H, Kotaja N (2011) Chromatoid body and small RNAs in male germ cells. Reproduction 142:195–209
119. Pascale A, Govoni S (2012) The complex world of post-transcriptional mechanisms: is their deregulation a common link for diseases? Focus on ELAV-like RNA-binding proteins. Cell Mol Life Sci 69:501–517
120. Ince-Dunn G, Okano HJ, Jensen KB et al (2012) Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. Neuron 75:1067–1080
121. Mukherjee N, Corcoran DL, Nusbaum JD et al (2011) Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. Mol Cell 43:327–339
122. Lebedeva S, Jens M, Theil K et al (2011) Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol Cell 43:340–352
123. Chi MN, Auriol J, Jegou B et al (2011) The RNA-binding protein ELAVL1/HuR is essential for mouse spermatogenesis, acting both at meiotic and postmeiotic stages. Mol Biol Cell 22:2875–2885
124. Vlasova-St Louis I, Dickson AM, Bohjanen PR, Wilusz CJ (1829) CELFish ways to modulate mRNA decay. Biochim Biophys Acta 2013:695–707
125. Dasgupta T, Ladd AN (2012) The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. Wiley Interdiscip Rev RNA 3:104–121
126. Vlasova-St Louis I, Bohjanen PR (2011) Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1. Curr Opin Genet Dev 21:444–451
127. Kress C, Gautier-Courteille C, Osborne HB, Babinet C, Paillard L (2007) Inactivation of CUG-BP1/CELF1 causes growth, viability, and spermatogenesis defects in mice. Mol Cell Biol 27:1146–1157
128. Sanchez-Jimenez F, Sanchez-Margalet V (2013) Role of Sam68 in post-transcriptional gene regulation. Int J Mol Sci 14:23402–23419
129. Iijima T, Wu K, Witte H et al (2011) SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. Cell 147:1601–1614
130. Missler M, Sudhof TC (1998) Neurexins: three genes and 1001 products. Trends Genet 14:20–26
131. Ehrmann I, Dalgliesh C, Liu Y et al (2013) The tissue-specific RNA-binding protein T-STAR controls regional splicing patterns of neurexin pre-mRNAs in the brain. PLoS Genet 9, e1003474
132. Paronetto MP, Zalfa F, Botti F, Geremia R, Bagni C, Sette C (2006) The nuclear RNA-binding protein Sam68 translocates to the cytoplasm and associates with the polysomes in mouse spermatocytes. Mol Biol Cell 17:14–24
133. Paronetto MP, Messina V, Bianchi E et al (2009) Sam68 regulates translation of target mRNAs in male germ cells, necessary for mouse spermatogenesis. J Cell Biol 185:235–249

134. Paronetto MP, Messina V, Barchi M, Geremia R, Richard S, Sette C (2011) Sam68 marks the transcriptionally active stages of spermatogenesis and modulates alternative splicing in male germ cells. Nucleic Acids Res 39:4961–4974
135. Sawicka K, Bushell M, Spriggs KA, Willis AE (2008) Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochem Soc Trans 36:641–647
136. Xu M, Hecht NB (2007) Polypyrimidine tract binding protein 2 stabilizes phosphoglycerate kinase 2 mRNA in murine male germ cells by binding to its 3′UTR. Biol Reprod 76:1025–1033
137. Schmid R, Grellscheid SN, Ehrmann I et al (2013) The splicing landscape is globally reprogrammed during male meiosis. Nucleic Acids Res 41:10170–10184
138. Licatalosi DD, Yano M, Fak JJ et al (2012) Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev 26:1626–1642
139. Wallace AM, Denison TL, Attaya EN, MacDonald CC (2004) Developmental distribution of the polyadenylation protein CstF-64 and the variant tauCstF-64 in mouse and rat testis. Biol Reprod 70:1080–1087
140. Yao C, Choi EA, Weng L et al (2013) Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3′ processing. RNA 19:1781–1790
141. Yao C, Biesinger J, Wan J et al (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proc Natl Acad Sci U S A 109:18773–18778
142. Martin G, Gruber AR, Keller W, Zavolan M (2012) Genome-wide analysis of pre-mRNA 3′ end processing reveals a decisive role of human cleavage factor I in the regulation of 3′ UTR length. Cell Rep 1:753–763
143. Berkovits BD, Wang L, Guarnieri P, Wolgemuth DJ (2012) The testis-specific double bromodomain-containing protein BRDT forms a complex with multiple spliceosome components and is required for mRNA splicing and 3′-UTR truncation in round spermatids. Nucleic Acids Res 40:7162–7175
144. Tardif S, Akrofi AS, Dass B, Hardy DM, MacDonald CC (2010) Infertility with impaired zona pellucida adhesion of spermatozoa from mice lacking TauCstF-64. Biol Reprod 83:464–472
145. Dass B, Tardif S, Park JY et al (2007) Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. Proc Natl Acad Sci U S A 104:20374–20379
146. Giorgini F, Davies HG, Braun RE (2002) Translational repression by MSY4 inhibits spermatid differentiation in mice. Development 129:3669–3679
147. Yang J, Medvedev S, Yu J, Schultz RM, Hecht NB (2006) Deletion of the DNA/RNA-binding protein MSY2 leads to post-meiotic arrest. Mol Cell Endocrinol 250:20–24
148. Yang J, Morales CR, Medvedev S, Schultz RM, Hecht NB (2007) In the absence of the mouse DNA/RNA-binding protein MSY2, messenger RNA instability leads to spermatogenic arrest. Biol Reprod 76:48–54
149. Tafuri SR, Familari M, Wolffe AP (1993) A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes. J Biol Chem 268:12213–12220
150. Mendez R, Murthy KG, Ryan K, Manley JL, Richter JD (2000) Phosphorylation of CPEB by Eg2 mediates the recruitment of CPSF into an active cytoplasmic polyadenylation complex. Mol Cell 6:1253–1259
151. Hosoda N, Funakoshi Y, Hirasawa M et al (2011) Anti-proliferative protein Tob negatively regulates CPEB3 target by recruiting Caf1 deadenylase. EMBO J 30:1311–1323
152. Pique M, Lopez JM, Foissac S, Guigo R, Mendez R (2008) A combinatorial code for CPE-mediated translational control. Cell 132:434–448
153. Campbell ZT, Menichelli E, Friend K et al (2012) Identification of a conserved interface between PUF and CPEB proteins. J Biol Chem 287:18854–18862
154. Kurihara Y, Tokuriki M, Myojin R et al (2003) CPEB2, a novel putative translational regulator in mouse haploid germ cells. Biol Reprod 69:261–268
155. Spassov DS, Jurecic R (2002) Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins. Gene 299:195–204

156. Zamore PD, Williamson JR, Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3:1421–1433
157. Xu EY, Chang R, Salmon NA, Reijo Pera RA (2007) A gene trap mutation of a murine homolog of the Drosophila stem cell factor Pumilio results in smaller testes but does not affect litter size or fertility. Mol Reprod Dev 74:912–921
158. Chen D, Zheng W, Lin A, Uyhazi K, Zhao H, Lin H (2012) Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis. Curr Biol 22:420–425
159. Suzuki A, Igarashi K, Aisaki K, Kanno J, Saga Y (2010) NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. Proc Natl Acad Sci U S A 107:3594–3599
160. Saga Y (2010) Function of Nanos2 in the male germ cell lineage in mice. Cell Mol Life Sci 67:3815–3822
161. Tsuda M, Sasaoka Y, Kiso M et al (2003) Conserved role of nanos proteins in germ cell development. Science 301:1239–1241
162. Beck AR, Miller IJ, Anderson P, Streuli M (1998) RNA-binding protein TIAR is essential for primordial germ cell development. Proc Natl Acad Sci U S A 95:2331–2336
163. Youngren KK, Coveney D, Peng X et al (2005) The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. Nature 435:360–364
164. Sada A, Suzuki A, Suzuki H, Saga Y (2009) The RNA-binding protein NANOS2 is required to maintain murine spermatogonial stem cells. Science 325:1394–1398
165. Reynolds N, Cooke HJ (2005) Role of the DAZ genes in male fertility. Reprod Biomed Online 10:72–80
166. VanGompel MJW, Xu EY (2011) The roles of the DAZ family in spermatogenesis: more than just translation? Spermatogenesis 1:36–46
167. Vogel T, Speed RM, Ross A, Cooke HJ (2002) Partial rescue of the Dazl knockout mouse by the human DAZL gene. Mol Hum Reprod 8:797–804
168. Saunders PT, Turner JM, Ruggiu M et al (2003) Absence of mDazl produces a final block on germ cell development at meiosis. Reproduction 126:589–597
169. Schrans-Stassen BH, Saunders PT, Cooke HJ, de Rooij DG (2001) Nature of the spermatogenic arrest in Dazl −/− mice. Biol Reprod 65:771–776
170. Ruggiu M, Speed R, Taggart M et al (1997) The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature 389:73–77
171. Lin Y, Page DC (2005) Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. Dev Biol 288:309–316
172. Reynolds N, Collier B, Bingham V, Gray NK, Cooke HJ (2007) Translation of the synaptonemal complex component Sycp3 is enhanced in vivo by the germ cell specific regulator Dazl. RNA 13:974–981
173. Reynolds N, Collier B, Maratou K et al (2005) Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. Hum Mol Genet 14:3899–3909
174. Xu X, Tan X, Lin Q, Schmidt B, Engel W, Pantakani DV (1829) Mouse Dazl and its novel splice variant functions in translational repression of target mRNAs in embryonic stem cells. Biochim Biophys Acta 2013:425–435
175. Tsui S, Dai T, Warren ST, Salido EC, Yen PH (2000) Association of the mouse infertility factor DAZL1 with actively translating polyribosomes. Biol Reprod 62:1655–1660
176. Kim B, Cooke HJ, Rhee K (2012) DAZL is essential for stress granule formation implicated in germ cell survival upon heat stress. Development 139:568–578

Chapter 7
Regulation of Stem Cell Self-Renewal and Oncogenesis by RNA-Binding Proteins

Ayuna Hattori*, Kristina Buac*, and Takahiro Ito

Abstract Throughout their life span, multicellular organisms rely on stem cell systems. During development pluripotent embryonic stem cells give rise to all cell types that make up the organism. After birth, tissue stem cells maintain properly functioning tissues and organs under homeostasis as well as promote regeneration after tissue damage or injury. Stem cells are capable of self-renewal, which is the ability to divide indefinitely while retaining the potential of differentiation into multiple cell types. The ability to self-renew, however, is a double-edged sword; the molecular mechanisms of self-renewal can be a target of malignant transformation driving tumor development and progression. Growing lines of evidence have shown that RNA-binding proteins (RBPs) play pivotal roles in the regulation of self-renewal by modulating metabolism of coding and non-coding RNAs both in normal tissues and in cancers. In this review, we discuss our current understanding of tissue stem cell systems and how RBPs regulate stem cell fates as well as how the regulatory functions of RBPs contribute to oncogenesis.

Keywords Tissue stem cells • Embryonic stem cells • Pluripotent stem cells • Tumor-initiating cells • Cancer stem cells • Self-renewal • Differentiation • Regeneration • Cancer development • Cancer progression • Oncogenesis • Heterogeneous ribonucleoprotein E2 • hnRNP E2 • Musashi • Lin28 • IGF2BP/IMP • HuR/Elav • EWS • TLS • eIF4E • PUM • C/EBPα • Wnt • Notch • Numb

1 Stem Cell Systems in Tissue Homeostasis and Oncogenesis

Stem cell biology represents a dynamic area of research because it provides great insights into basic cellular processes including development, aging and oncogenesis and as such holds vast clinical implications and therapeutic potential for multiple

*These authors contributed equally to this work.

A. Hattori • K. Buac • T. Ito (✉)
Department of Biochemistry and Molecular Biology, University of Georgia,
500 D.W. Brooks Drive, Athens, GA 30602, USA
e-mail: ito@bmb.uga.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine
and Biology 907, DOI 10.1007/978-3-319-29073-7_7

diseases in the era of regenerative medicine [1]. The main characteristic of stem cells is their ability to propagate for indefinite periods while retaining cellular potential to differentiate into multiple cell types, a process called self-renewal. Based on their potency, several types of stem cells have been described: embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), germline stem cells (GSCs) and adult/tissue stem cells. ESCs, iPSCs and GSCs are pluripotent cells that have the ability to form any of the cell lineages in the body and are functionally similar to one another [2, 3]. Adult or tissue stem cells are multipotent and can generate all cell types of a given tissue or lineage. The role of tissue stem cells is to maintain tissue function and architecture by replenishing short-lived mature cells during homeostasis and injury. Tissue stem cells often occur at low frequency and have been identified in several organs and tissues such as the blood, digestive tract, mammary gland, brain, skin and skeletal muscle. To illustrate the concept of stem cell self-renewal and differentiation as well as their hierarchical ontology, in the following sections we discuss our current knowledge on the hematopoietic and intestinal stem cell systems in more detail.

1.1 Hematopoietic Stem Cells

The mammalian blood lineage is composed of several distinct cell types including erythrocytes, megakaryocytes/platelets, monocytes/macrophages, granulocytes, mast cells, T- and B- lymphocytes, natural killer cells and others. All these cells are derived from a single type of cell termed hematopoietic stem cells (HSCs) [4]. HSCs reside within bone marrow, exist at very low frequency (0.01 % of the total nucleated cells) and replenish the mature blood cell pool during homeostasis and stress through an orchestrated process of self-renewal and differentiation [5–8]. Similar to other tissue stem cells, HSCs can asymmetrically divide to produce transit-amplifying progenitors in a hierarchical fashion (Fig. 7.1). The long-term HSCs give rise to short-term HSCs, which in turn give rise to multipotent progenitors (MPPs) without detectable self-renewal potential [9]. MPPs generate all the mature myeloid, lymphoid and erythroid precursor cells through a series of cell fate commitment steps unique to each cell lineage [10–12]. At the bench, HSCs can be purified from bone marrow based on their cell surface phenotype using fluorescence-activated cell sorting (FACS) techniques. HSCs are negative for lineage-specific markers (such as B220, CD4, CD8, Gr-1, CD11b/Mac-1 and TER-119) and positive for Sca-1 and c-Kit cell surface markers; hence these are commonly referred to as the LSK population (for Lin$^-$ Sca-1$^+$ c-Kit$^+$) [13]. The cell population defined by the LSK surface phenotype contain both MPPs and HSCs, and utilization of additional cell surface markers, such as Thy1.1, CD34, CD150, Flk2/Flt3 and endoglin or use of the Hoechst 33342 dye-efflux profile, can be employed to further enrich HSCs [8, 9, 14–18]. Upon transplantation into a recipient animal, a single purified murine HSC can engraft and repopulate the recipient's bone marrow and reconstitute the entire blood population through its ability of self-renewal and multi-lineage differentiation [4].

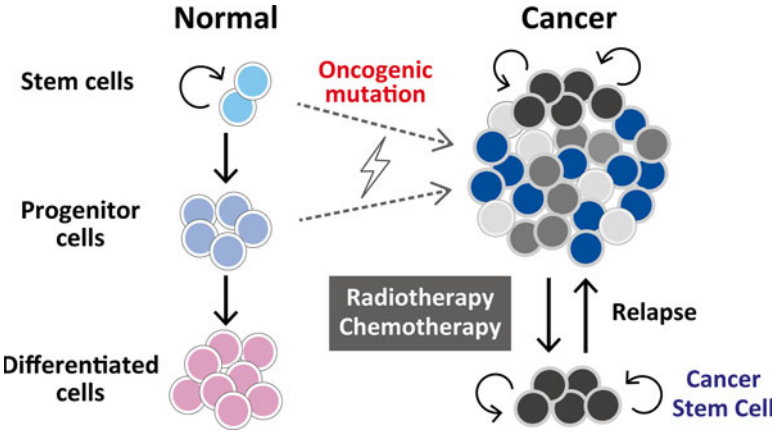


Fig. 7.1 Stem cell systems in normal tissues and cancer. Tissue stem cells are capable of differentiation and self-renewal. Self-renewal is the process by which stem cells can generate more cells while maintaining the undifferentiated state. As undifferentiated cells become committed to differentiate, they progressively lose the self-renewing potential. These two fundamental abilities of stem cells are essential for morphological and functional tissue architecture of multiple tissues such as hematopoietic, mammary, colorectal or neural tissues in metazoa. Tumors are phenotypically and functionally heterogeneous, and they often share the similar hierarchical organization as in normal tissues. Because some of the tumor cells are capable of self-renewal and often therapy-resistant due to their quiescent nature, these cells can "regenerate" tumors after therapy cessation causing tumor relapse or secondary tumors via metastasis. The self-renewing cancer cells, or cancer stem cells, are likely being generated via aberrant activation of self-renewal program by oncogenic mutations in normal stem or progenitor cells

1.2 Intestinal Stem Cells

In the intestine, stem cells replenish the epithelium of both the small and large intestine every 4–5 days. The intestinal epithelium is composed of villi, which are luminal protrusions, and crypts, which are pit-like recessions. Intestinal stem cells reside at the bottom of crypt [19–21]. As intestinal stem cells divide, their daughter cells exit the stem cell compartment and move into the transit-amplifying area, where they go through 4–5 divisions. During this process, the transit-amplifying cells move towards the crypt-villus junction and differentiate into several types of intestinal cells such as enterocytes, goblet cells, Paneth cell and enteroendocrine cells. As they differentiate, cells migrate upwards towards the tip of the villus and they are shed into the gut lumen after maturation. Paneth cells are an exception to this upward migration as they migrate down into the base of the crypt and remain there for up to 6–8 weeks [22]. To date two types of intestinal stem cells have been described that reside in the crypt of small intestine: *Lgr5*-positive, slow cycling crypt base columnar cells, and quiescent, radiation-resistant '+4' cells marked by the expression of *Bmi1*, *Hopx*, *Tert*, and *Lrig1* [23–27]. $Lgr5^{+}$ cell can self-renew

and differentiate into all intestinal epithelium cell types, and '+4' cells can give rise to Lgr5$^+$ cells, indicating that both have stem cell activities [28–30]. The major distinction between the two population is that Lgr5$^+$ stem cells are mitotically active than '+4' stem cells. Together these data suggest that Lgr5$^+$ cells are responsible for maintaining intestinal homeostasis while '+4' cells function as injury-inducible reserve stem cells in vivo [26, 30].

1.3 Cancer Stem Cells

Tumors, like normal tissues, are composed of various types of cells that exhibit varying capacities of self-renewal and differentiation (intratumor heterogeneity). There are two major hypotheses as to how the cellular heterogeneity is generated. In the stochastic model, all tumor cells are functionally equivalent, with intrinsic and extrinsic factors contributing stochastically to select tumor cells that initiate tumor growth. In the hierarchy model, a fraction of tumor cells, called tumor-initiating cells (TICs) or cancer stem cells (CSCs), propagate the tumor (Fig. 7.1) [31–34]. Studies in the past decade have shown that many tumors, if not all, follow the hierarchy model [35–38]. CSCs exhibit distinct cell surface phenotype and therefore can be prospectively isolated by cell sorting. Following transplantation, CSCs can regenerate heterogeneous populations of tumor cells because of their potential to self-renew and differentiate. Cell surface markers defining the CSC population have been identified in several types of cancers. For example, in human acute myeloid leukemia, a population with CD34$^+$CD38$^-$ surface phenotype, can engraft immune-deficient SCID mice while CD34$^+$CD38$^+$ or CD34$^-$ cells do not [35, 39]. In human colorectal cancer, Prominin/CD133 or the combination of EpCAM and CD44 can separate the cells with tumorigenic potential from those without [40–42]. While CSCs are the self-renewing population within a tumor, they do not necessarily originate from normal stem cells in the corresponding tissue [43]. In fact, CSCs can arise independent of tissue stem cells. For example, in acute myeloid leukemia induced by the MLL-AF9 gene fusion, multiple studies with experimental cancer models have clearly demonstrated that committed myeloid progenitors can be transformed into CSCs in this type of leukemia [44–46].

Since CSCs, but not the non-CSCs, self-renew to sustain the tumor for long periods of time, the tumor's characteristics are likely determined by the properties of the CSC population in the tumor. For example, in medulloblastoma, there are four major tumor types known with each type arising from distinct populations of either stem or progenitor cells [47, 48]. Also, CSCs often represent a population resistant to conventional therapeutic interventions such as radiation or chemotherapy and are therefore responsible for relapse after therapy cessation [49–51]. The ability of CSCs to initiate a new tumor implies that the population is also responsible for the formation of remote secondary tumors via metastasis. Therefore, it is critical to understand the cellular properties of CSCs and molecular regulation of their self-renewal for establishing effective targeted therapies.

2 Developmental Signals Regulating Stem Cell Self-Renewal

Multiple signaling pathways are involved in regulation and maintenance of normal stem cell populations. Frequently, these pathways are shared by CSCs in propagating and sustaining a tumor. To date, several developmental pathways operating during embryogenesis are implicated in both stem cell biology and oncogenesis, which include Wingless/int (Wnt) and Notch among others.

Wnt signaling plays a role both in development and cancer progression [52]. Wnt ligands are secreted glycoproteins that bind to cell surface receptors and activate a multitude of molecular targets within a cell [53–55]. Their ability to activate multiple signaling pathways is in part mediated by a vast number of ligands (19 ligands identified in mammals alone) and multiple receptors (encoded by the Frizzled and LRP genes) that are expressed in a distinct spatiotemporal manner during embryo development and in adulthood [56]. The most studied canonical Wnt pathway is mediated by β-catenin and leads to changes in gene expression that influence cell fate choice [57, 58]. In the absence of Wnt ligand, the degradation complex composed of Glycogen Synthase Kinase 3 (GSK3), Axin/Conductin, Casein kinase 1 (CK1) and Adenomatous Polyposis Coli (APC) gene product sequesters β-catenin from the cytoplasm. Within this complex β-catenin is constitutively phosphorylated by CK1 and GSK3, and phospho-β-catenin is recognized by the E3 ubiquitin ligase βTrCP/Fbwx1, which targets β-catenin for ubiquitin-dependent proteasomal degradation. Binding of Wnt ligand to its receptor activates Dishevelled via phosphorylation, which in turn recruits Axin and the degradation complex to the membrane that prevents β-catenin degradation by the proteasome. This leads to an increase in cytoplasmic levels of β-catenin which then translocates into the nucleus, where it binds to a TCF/LEF transcription factor that activates downstream target genes such as *Ccnd1* (Cyclin D1). The Wnt signaling pathway has been implicated in specification and maintenance of stem and progenitor cells in various tissues and organs. For instance, targeted deletion of Tcf4 leads to complete ablation of intestinal epithelial stem cells and postnatal lethality in mice [59]. Later studies have shown that canonical Wnt signaling interacts with BMP and Notch signals in the intestinal microenvironment to control stem cell self-renewal [60, 61]. Another example is Sox9 regulation by the Tcf4-β-catenin complex in the intestinal epithelial cells [62, 63]. Sox9 is a transcription factor expressed in intestinal stem/progenitor and in Paneth cells, and conditional deletion results in an increased differentiation throughout small intestine and a complete loss of Paneth cells, suggesting that the Wnt-Tcf4-Sox9 axis is required for proper maintenance of intestinal stem cells. The canonical Wnt pathway has also been implicated in the self-renewal and differentiation of HSCs as well. Reya *et al.* have demonstrated that overexpression of a constitutively active β-catenin in HSCs of *H2K-BCL2* transgenic mice results in a significant increase in stem cell numbers while maintaining their multipotency [64]. In mouse ESCs, Wnt signaling through Tcf3 promotes self-renewal and suppresses differentiation by transcriptionally activating core pluripotency factors Sox2, Oct4, and Nanog [65]. A recent study has shown that Tcf3

downregulation by Wnt-inducible *miR-211* promotes self-renewal and prevents neural differentiation in mouse ESCs [66]. Collectively, these findings highlight the wide contributions of Wnt signaling to the regulation of stem cell fates.

Another developmental pathway shown to regulate self-renewal and oncogenesis is Notch signaling. This pathway plays a critical role in the development and patterning of a wide variety of organisms including worms, flies, and mice [67]. Binding of Notch ligands (encoded by Delta-like or Jagged genes) activates the Notch receptor by proteolytic cleavage releasing the Notch intracellular domain (NICD) from the cell membrane and its translocation into the nucleus. Once inside the nucleus, NICD associates with DNA binding proteins of the CSL family transcriptional regulator (RBPJ in human and mice) to activate target gene transcription. One of the many Notch target genes are the basic helix-loop-helix transcriptional repressors of the Hairy/Enhancer of Split (Hes) and Hes-related (Hey) proteins that suppress expression of downstream targets, such as tissue specific transcription factors [68, 69]. The role of Notch signaling in stem cell maintenance has been implicated in multiple tissues. Conditional deletion of *Rbpj* in the myogenic lineage lead to complete depletion of the $Pax3^{+}$ $Pax7^{+}$ muscle progenitor cells accompanied by an increase in the $MyoD^{+}$ differentiating myoblasts [70]. This resulted in a postnatal skeletal muscle defect that was rescued by loss of *MyoD*, suggesting that Notch regulation of muscle progenitor cells is mediated, in part, by *MyoD* suppression [71]. Activation of Notch signaling in mesodermal cells prevents the formation of cardiac muscle, endothelial, and hematopoietic cells while inhibition of Notch signaling in ESCs promotes differentiation along the cardiomyocyte lineage [72]. Furthermore, Notch activation increases HSC self-renewal both in vivo and in vitro [73–75]. In contrast, activated Notch signal promotes neural differentiation in ESCs and airway basal stem cells in the lungs [76, 77]. These data suggests that the regulatory effect of Notch signaling on self-renewal and differentiation is both cell type-specific and context-dependent.

3 RNA-binding Proteins: The Emerging Players in Stem Cell Regulatory Network

Based on multiple studies from different groups, RNA binding proteins (RBPs) constitute an additional layer of fine-tuning of stem cell self-renewal and cell fate determination during development and oncogenesis. RBPs bind to their target RNAs in different cellular compartments and at different steps of RNA metabolism therefore providing spatiotemporal control of gene expression. Processes regulated by RBPs include mRNA capping, splicing, cleavage, nontemplated nucleotide addition, nucleotide editing, nuclear export, intracellular localization, stability, and translation [78, 79]. Most RBPs associate with their target mRNAs through specific nucleotide sequences located in the untranslated regions (UTRs) as well as in the coding or open reading frame (ORF) regions of the transcript. The specific binding is achieved through RNA-binding domains (RBDs). Currently more than 40 RBDs

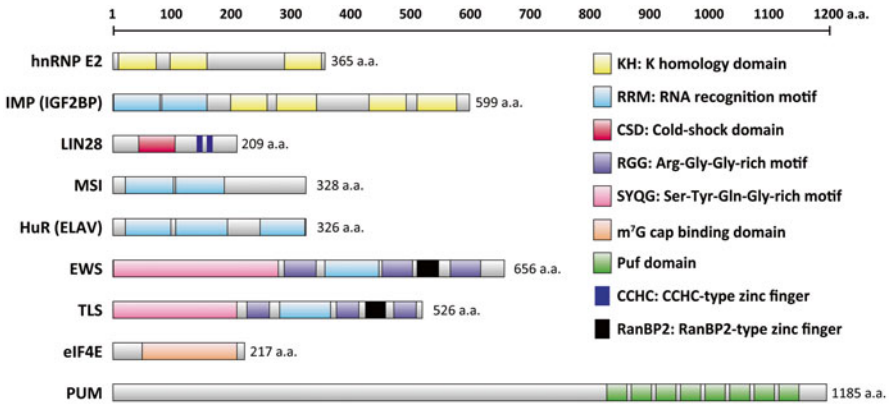


Fig. 7.2 Domain structures of the RBPs discussed. Schematics of the eight RNA binding proteins we discuss in this article. **hnRNP E2** contains the K-homology (KH) domain which is an evolutionary-conserved RNA-binding domain. **MSI** and **HuR/ELAV** contain several RNA Recognition Motifs (RRMs). **IMP/IGF2BP** and **LIN28** harbors two distinct types of RNA-binding domains, RRMs and KH domains, and a Cold Shock Domain (CSD) and CCHC zinc fingers (ZnFs), respectively. **EWS** and **TLS** proteins of FET family share similar structure that contains an RRM, RGG motifs and RanBP2 type ZnF, and the SYQG motif at the N-terminus function as a transcriptional activation domain. **eIF4E** is an mRNA cap-binding protein. **PUM** proteins contain a C-terminal Puf domain that consists of eight Puf repeats consisting of ~40 amino acids

have been identified and are commonly used to classify RBPs into different families. In this section we discuss the functions of eight RBP families that have been implicated in regulation of stem cell self-renewal and oncogenesis (Fig. 7.2).

3.1 Heterogeneous Ribonucleoprotein E2 (hnRNP E2)

The RNA-binding protein hnRNP E2, also known as poly(rC) binding protein (PCBP2), is one of the best studied RBPs with a K Homology (KH) domain. As many other hnRNP proteins do, hnRNP E2 shuttles between the cytoplasm and the nucleus regulating pre-mRNA processing, mRNA localization, translation and stability. Initially identified in the hnRNP K protein, the KH domain is an evolutionarily-conserved RNA-binding domain consisting of 65–70 amino acids. HnRNP E2 binds to C-rich sequences via the KH domain and regulates cell differentiation status in several types of tumors (tumor grades) [80–83].

In chronic myeloid leukemia (CML), for example, hnRNP E2 regulates disease progression by inhibiting C/EBPα translation (Fig. 7.3). CML is a myeloproliferative disorder initiated by the BCR-ABL chromosomal translocation and is characterized by distinct disease phases, namely, chronic phase (CP) and blast crisis phase (BC) [84]. In CP-CML, patients have increased myeloid cell counts and mature granulocytes are still produced. As the disease progresses to BC-CML, hematopoietic differentiation becomes arrested and immature myeloid progenitors expand and

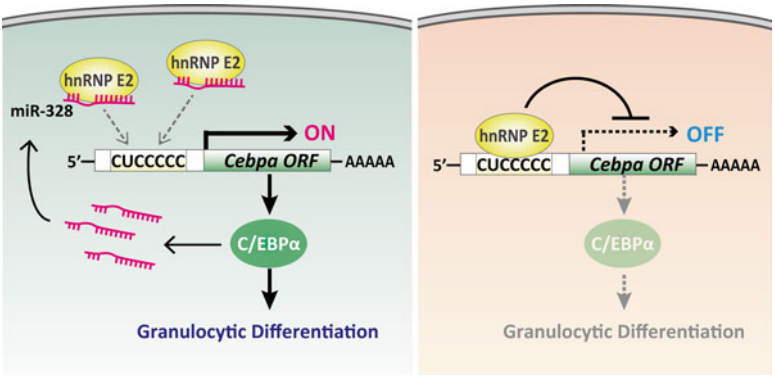


Fig. 7.3 hnRNP E2 repression of C/EBPα protein expression. In self-renewing cells, hnRNP E2 recognizes and directly binds to the sequence CUCCCCC within the 5′-UTR of *Cebpa* transcripts, which directly inhibits their translation (*right panel*). In differentiating cells, *miR-328*, whose sequence is similar to the hnRNP E2 binding element, is expressed and function as a decoy. The binding of *miR-328* to hnRNP E2 protein relieves *Cebpa* transcripts from translational inhibition, and C/EBPα induces granulocyte differentiation. In addition, C/EBPα activates *miR-328* transcription, forming a positive feedback loop stabilizing the granulocytic differentiation signal

accumulate in the bone marrow. Because BC-CML is often refractory to current therapies and is therefore associated with poor prognosis, how the transition from indolent CP-CML to more aggressive BC-CML occurs is one of the critical problems to be solved. During CML disease progression, hnRNP E2 protein becomes highly upregulated and inhibits the translation of C/EBPα mRNA [82]. The transcription factor C/EBPα is an essential regulator of granulocytic differentiation by activating expression of several key factors in myeloid maturation, including granulocyte colony-stimulating factor (G-CSF) receptor. The 5′-UTR of the *C/EBP*α mRNA contains a conserved C-rich element, CUCCCCC. By directly binding to the C-rich element, hnRNP E2 inhibits *C/EBP*α mRNA translation, thereby suppressing the myeloid differentiation program, which contributes to retain cancer stem cell activity in BC-CML. The aberrant upregulation of hnRNP E2 protein in BC-CML is mediated by phosphorylation via BCR-mediated MAPK activation. ERK1/2 MAPK phosphorylates hnRNP E2 at Ser-173, 189, 272 and Thr-213, and these phosphorylations stabilize the hnRNP E2 protein [85]. In more recent studies, Eiring *et al.* demonstrated that *miR-328* counteracts hnRNP E2 in translational repression of *C/EBP*α mRNA and induces differentiation of immature BC-CML cells into mature granulocytes [86]. *miR-328* directly interacts with hnRNP E2 protein and competitively inhibits hnRNP E2 binding to *C/EBP*α mRNA, indicating that the miRNA acts as a molecular decoy for hnRNP E2. Forced *miR-328* expression rescues hematopoietic maturation of the differentiation-arrested cells through restoration of *C/EBP*α mRNA translation. As a result, *miR-328* successfully impairs the clonogenic potential of human BC-CML cells in vitro and reduces tumor burden in mouse xenograft model, suggesting that the hnRNP E2-C/EBPα axis is important for leukemia stem cell survival in vivo. The authors show not only that *miR-328*

regulates C/EBPα expression but also that the C/EBPα protein itself induces *miR-328* expression by binding directly to the miR-328 gene promoter, which highlights the complexity of the hnRNP E2 regulatory network.

The regulatory role of hnRNP E2/PCBP2 in cell differentiation seems conserved in other cancers. Recently hnRNP E2 was found to be highly expressed in glioma [80]. Gene and protein expression analysis on primary samples from glioma patients show that its expression level is correlated with the tumor grade, *i.e.*, elevated hnRNP E2 levels in poorly-differentiated tumors (higher grade tumors). Short hairpin RNA (shRNA)-mediated knockdown increased cell death in human glioma cell lines and reduced intracranial tumor growth after xenograft, implying its role in glioma stem cell renewal in vivo. RNA-immunoprecipitation analysis (RIP) of this RNA-binding protein identified 35 potential mRNA targets. Among those, *FHL3* is down-regulated in human glioma. Through interaction with the C-rich element in *FHL3* mRNAs, hnRNP E2 inhibits *FHL3* translation and thus blocks FHL3-induced apoptosis, providing a molecular mechanism how this RNA-binding protein contributes to glioma malignancy.

In contrast, in oral squamous cell carcinomas, hnRNP E2 expression is higher in well-differentiated than poorly-differentiated tumors [83]. Functionally, overexpression of hnRNP E2 in the carcinoma cells results in reduced growth and increased apoptosis. These reports imply that oncogenic roles of hnRNP E2 are context- and tumor-dependent. Identification of hnRNP E2 target RNAs in each tumor may help us to understand the distinct roles of the RBP in malignant self-renewal and cell differentiation.

3.2 IGF2BP/IMP Family

The insulin-like growth factor-2 mRNA-binding proteins (IGF2BPs) are members of conserved family of RNA-binding proteins that contain two N-terminal RNA Recognition Motifs (RRMs) and four C-terminal KH domains [87]. Three paralogues are known in mammals, namely IGF2BP1, IGF2BP2, and IGF2BP3. Nomenclature of IGF2BPs has gone through many rounds of evolution and in the literature synonyms used to describe this protein family include: IMP, CRD-BP, VICKZ, ZBP, Vg1RBP/Vera or KOC. The IGF2BPs exert wide array of biological functions and have been implicated in post-transcriptional processes such as mRNA localization, transport, turnover, and translational control [88].

All three proteins are “oncofetal”, meaning they are silenced or expressed at very low levels in the adult tissue while they are highly expressed during early embryonic development and expressed *de novo* in malignancies. For example, IGF2BP1/IMP1 was originally identified as a stabilizing factor for several mRNAs such as *c-Myc*, *β-actin* and *IGF-II* [89–92]. During development *Igf2bp1* is ubiquitously expressed in mice, and its loss of function results in growth retardation at late gestation and high perinatal mortality [91, 93]. After birth, these mice exhibit impaired development of the small intestine and colon. In adulthood, the gene is expressed in very

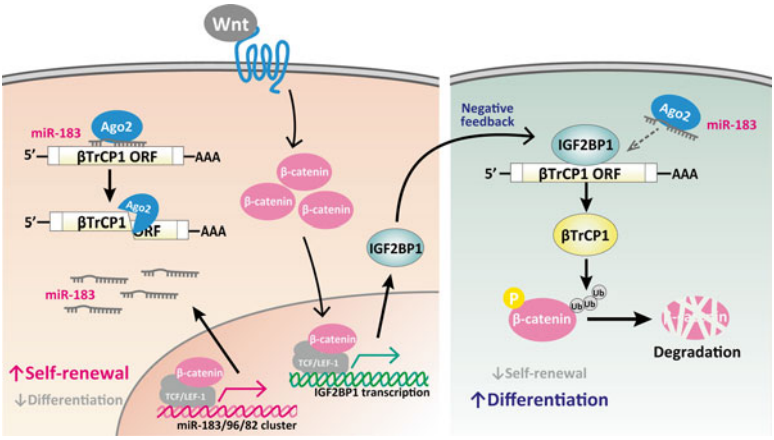


Fig. 7.4 IGF2BP1 regulates canonical Wnt signaling pathway. In Wnt-responsive cells, activated β-catenin and TCF/LEF-1 upregulate *miR-183* expression. In the absence of IGF2BP1, *βTrCP1* mRNAs are constitutively degraded via *miR-183*-dependent, Ago2-mediated endonucleolytic cleavage. IGF2BP1 binding to *βTrCP1* mRNAs prevents *miR-183*-dependent degradation, leading to stabilization of the mRNA and increased βTrCP1 protein expression. Elevated levels of βTrCP1 results in increased SCFβ^{TrCP1} E3 ubiquitin ligase activity and accelerate β-catenin degradation

few tissues, suggesting that this factor predominantly regulates development and morphogenesis [93, 94].

In contrast, many human cancers show highly activated *IGF2BP1* expression including melanoma, colon, lung, skin, testicular, mammary and ovarian tumors [95–100]. Knockdown of *IGF2BP1* in MCF-7 breast cancer cells impaired proliferation and downregulated tumorigenic genes such as *c-MYC*, highlighting its role in tumor maintenance [101]. In colorectal carcinoma (CRC), patients with IGF2BP1-positive tumors show higher incidence of metastasis and poorer prognosis, identifying IGF2BP1 as a potential diagnostic marker for CRC [95]. Consequently *IGF2BP1* knockout resulted in attenuated intestinal tumorigenesis while overexpression led to enhanced metastasis and increased number of circulating tumor cells, implying that IGF2BP1 plays an essential role in CSC self-renewal [102]. IGF2BP1 protein augments K-Ras expression by direct binding to and stabilization of *K-Ras* mRNA, which is an important driver of CRC [103, 104]. Taken together, aberrant *IGF2BP1* activation seems to mediate several aspects of CRC development. IGF2BP1 also regulates Wnt/β-catenin/Tcf signaling by modulating the stability of *βTrCP1* mRNA (Fig. 7.4). As described earlier in this chapter, βTrCP1 encodes a subunit for the SCF E3 ubiquitin ligase, which targets β-catenin protein for degradation. *Mir-183*, a Wnt/β-catenin/Tcf target microRNA gene, promotes degradation of *βTrCP1* mRNAs by the Ago2-mediated endonucleolytic cleavage in the absence of IGF2BP1 protein [105, 106]. IGF2BP1 binding to *βTrCP1* mRNA prevents *miR-183*-dependent degradation, leading to stabilization of the mRNA and elevated βTrCP1 protein translation and expression. Since IGF2BP1 expression itself is activated by β-catenin, IGF2BP1-mediated stabilization of *βTrCP1* mRNA constitutes

a negative feedback regulatory loop in the β-catenin signals, and thus contributes to the regulation of Wnt-dependent stem cell maintenance in vivo.

IGF2BP2, similar to IGF2BP1, is expressed during early development but its expression ceases in most adult tissues [91, 107, 108]. In mouse neocortex, *Igf2bp2* regulates differentiation potential of neural precursor cells (NPCs) in developing mouse brain [109]. *Igf2bp2* is highly expressed in the early-stage NPCs but reduced in the late-stage NPCs and absent in adult neurons. Functionally, high *Igf2bp2* expression mediates the neurogenic potential of the early-stage NPCs, and ectopic *Igf2bp2* expression in the late-stage NPCs increases neurogenesis at the expense of glial differentiation. *Igf2bp2* appears to regulate the neurogenic-to-gliogenic transition of NPCs and thus function as a key regulator of cell fates during development.

Several reports indicate that *IGF2BP2* expression is also re-activated in human malignancies [97, 110, 111]. In grade IV glioblastoma (GBM), the most aggressive form of brain tumor and one of the most refractory malignancies to current clinical therapies, the IGF2BP2 expression is elevated while lower grade tumors (grades II and III) have significantly lower levels [110]. Within GBM, *IGF2BP2* is higher in $CD133^+$ self-renewing tumor cells, that is, in CSCs. *IGF2BP2* knockdown leads to reduced gliomasphere formation and size in vitro. When these cells are injected into NOD/SCID mice their survival was extended compared to mice injected with control cells. Thus loss-of-function experiments indicate that *IGF2BP2* is necessary for glioma stem cell self-renewal and proliferation. RIP analysis identified IGF2BP2 association with multiple cytochrome c oxidase transcripts encoding mitochondrial respiratory chain complex IV (CIV) subunits essential for oxidative phosphorylation. In addition, IGF2BP2 directly interacts with the mitochondrial respiratory chain complex I (CI) protein and controls its assembly. Accordingly, loss of *IGF2BP2* impairs the CI and CIV functions, leading to decrease in oxygen consumption and loss of clonogenic gliomasphere forming activity. Thus, one of previously unrecognized functions of IGF2BP2 is the regulation of assembly and function of mitochondrial respiratory chain, which further promotes self-renewal of glioma stem cells.

Like the other two members, IGF2BP3 is also an oncofetal protein, but its role in cell physiology has been poorly characterized [112]. It is expressed in human ESCs and becomes activated in multiple malignancies such as colorectal, pancreatic and ovarian cancers. Its high expression is often associated with poor prognosis [113–115]. One study on hepatocellular carcinoma highlights the importance of IGF2BP3 in regulating tumorigenic activity of TICs/CSCs [116]. TLR4 activation in $CD133^+$ CSCs promotes tumorigenesis by inducing *Nanog* expression. Nanog then activates expression of *Igf2bp3* and *Yap*, both of which have an ability to attenuate the TGF-β pathway by preventing Smad3 phosphorylation and its nuclear translocation. Igf2bp3-mediated activation of IGF-II stimulates the Akt/mTOR pathway that suppresses Smad3 phosphorylation while Yap1 interaction with Smad3 promotes cytoplasmic retention [117, 118]. Silencing of both *Igf2bp3* and *Yap* expressions resulted in the accumulation of phosphorylated nuclear Smad3 and delayed tumor growth with concomitant reductions of stem cell gene expression (*Nanog, Cd133, Oct4* and *Sox2*), suggesting that *Igf2bp3* and *Yap* regulate cancer stem cell propagation via the TGF-β signaling pathway. A role of IGF2BP3 in

CSCs is further supported by studies on osteosarcoma, a high-grade malignant bone tumor [119, 120]. In this malignancy, *IGF2BP3* expression is activated during tumor formation, and this increase is correlated with transformation of osteosarcoma cells into more aggressive CSCs that can metastasize. Consequently, knockdown of *IGF2BP3* in osteosarcoma cell lines inhibited tumor growth and suppressed metastasis, suggesting that elevated *IGF2BP3* confers oncogenic potential. These results together with findings from hepatocellular carcinoma implicate IGF2BP3 as a critical regulator of CSCs and as such represents a potential therapeutic target for these malignancies.

3.3 Lin28

RBPs not only regulate mRNA but also mediate regulation of miRNA biogenesis. The developmentally regulated RNA-binding protein Lin28 was initially identified in *C. elegans*, and many studies to date have demonstrated that the factor associates with *let-7* miRNA precursors and inhibits their processing and maturation [121–123]. There are two paralogous genes, *Lin28a* and *Lin28b*, encoding proteins that harbor two types of RNA-binding domains, a Cold Shock Domain (CSD) and two CCHC zinc fingers. Both Lin28A and B appear to function interchangeably although only Lin28B contains both a nuclear localization signal (NLS) and a nucleolar localization signal (NoLS), which account for its nucleolar localization. In *let-7* biogenesis (Fig. 7.5), Lin28 protein competes with the miRNA machineries for *let-7* binding during both steps of processing by Drosha (pri-*let-7*) and Dicer (pre-*let-7*) [124–126].

The functional significance of the Lin28-let-7 axis in stem cell renewal has been well characterized in ESCs and iPSCs [127, 128]. Lin28 is highly expressed in undifferentiated ESCs and thus the level of mature *let-7* miRNAs is kept low. In ESCs and iPSCs, the pluripotency transcription factors such as Oct4, Sox2 and Nanog occupy the *Lin28* gene promoter and positively regulate its gene transcription [129]. As ESCs differentiate, Lin28 expression declines and *let-7* in turn becomes more abundant in differentiated cells. Accordingly, forced expression of *let-7* is sufficient to functionally rescue the defective differentiation phenotype of *Dgcr8*-deficient ESCs, which lack miRNAs [130]. Interestingly, *Lin28* mRNA has *let-7* binding elements in its 3′-UTR, indicating that *Lin28* is a direct target of *let-7* miRNA. *let-7* also targets other pluripotency factors including c-Myc and Sall4, and functional inhibition of *let-7* facilitates reprogramming of fibroblasts to iPSCs, indicating that Lin28-mediated repression of *let-7* is critical in maintaining stem cell fate in ESCs and iPSCs [129] (Fig. 7.5).

LIN28 is overexpressed in a broad range of human cancers [131, 132]. The first definitive evidence of its functional role was reported in breast cancer. *let-7* expression was found to be low in the CSC population of this cancer. Forced expression of *let-7* induces cell differentiation and results in reduced ability in mammosphere formation in vitro as well as tumor-initiating CSC activity in vivo, indicating a repressive role for *let-7* in tumor self-renewal [133]. Moreover, *let-7* silencing by

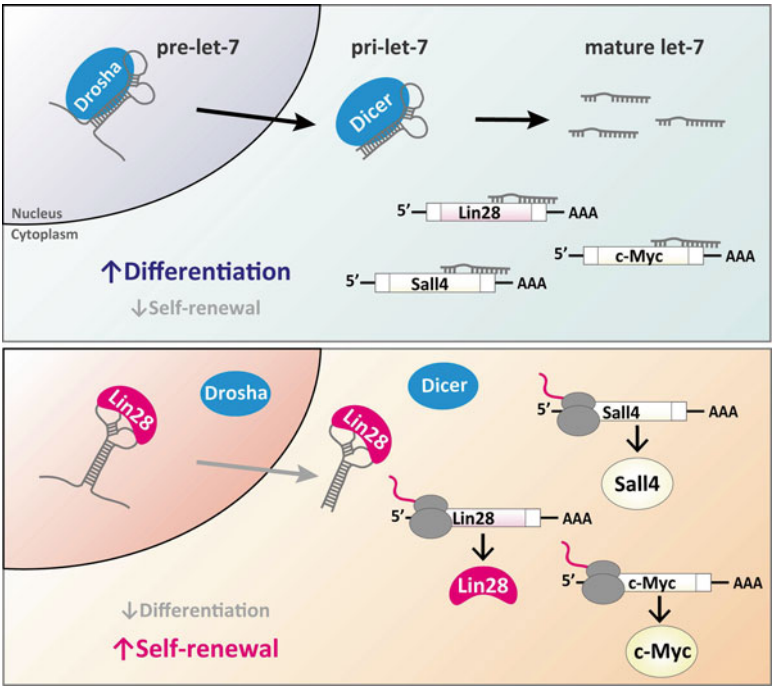


Fig. 7.5 Lin28 regulation of *let-7* biogenesis and stem cell differentiation. In differentiated cells (*top panel*), Drosha and Dicer complexes process pre- and pri-*let-7* transcripts, respectively, generating mature microRNA. Mature *let-7* RNA binds to and inhibits translation of mRNAs encoding "stem cell factors" such as c-Myc and Sall4. In undifferentiated cells, Lin28 binds to terminal loop structures of pre-*let-7* and antagonize cleavage and processing, which result in an increased expression of its targets, including Lin28 itself

antisense oligonucleotides enhanced sphere-forming ability in SK-BR-3 breast cancer cells, and antisense treatment also increased $CD44^{high}CD24^{low}$ cells that have CSC activity, suggesting that *let-7* down-regulation can reprogram differentiated tumor cells into tumor-initiating stem cells. The Wnt/β-catenin pathway mediates direct activation of *Lin28* transcription, which in turn represses *let-7* and induces CSC expansion [134]. Another critical regulator of *Lin28* gene expression is the transcription factor Dachshund (Dach1) [135]. Dach1 directly binds to the promoter region of *Lin28*, *Sox2* and *Nanog* genes and negatively regulates their expression in breast cancer cells. Consistent with this repressive role, Dach1 expression significantly inhibits expansion of $CD44^{high}CD24^{low}$ CSC population and therefore blocks breast tumor growth in vivo. LIN28 also plays a role in cancer metastasis [136]. Gene knockdown experiments clearly showed that bone metastasis of the MDA-MB-231-1833 breast tumor cells is highly dependent on Lin28 expression via its ability to repress *let-7* expression. Together, these and other studies established an essential role of Lin28-let-7 axis in CSC self-renewal, maintenance and metastasis of mammary tumors.

Lin28 has also been shown to regulate both normal and malignant self-renewal during hematopoiesis. Upon transplantation, fetal HSCs expand faster and achieve higher levels of engraftment than adult counterparts do. Copley *et al.* found that Lin28 is abundantly expressed and de-represses *let-7*-mediated *Hmga2* repression in fetal HSCs [137]. Ectopic expression of *Lin28* in purified adult HSCs upregulates *Hmga2* and augments their self-renewal activity. Since *Hmga2*-deficient fetal HSCs exhibit impaired self-renewal activity, the high self-renewing ability of fetal HSCs is likely to be mediated by the *Lin28-let-7-Hmga2* axis. In CML, *Lin28* expression is more frequent in BC-CML specimens than in CP-CML [132]. Inhibiting *Lin28* expression in a BC- CML cell line K562 restores the levels of mature *let-7* miRNA and impairs the proliferative and clonogenic capacities with concomitant induction of cellular differentiation, suggesting that silencing of Lin28 might delay or block the aggressive transformation to the blast crisis phase. Considering the robust ability of Lin28 to facilitate self-renewal in multiple types of normal and cancer stem cells, modulating the Lin28 pathway could prove an effective strategy in controlling pluripotency and treating human cancer.

3.4 Musashi Family

The first member of the Musashi (Msi) family of RNA-binding proteins was discovered in *Drosophila* as a key regulator of asymmetric cell division, stem cell function and cell fate determination [138, 139]. To date, *Msi* genes have been identified in both vertebrates and invertebrates, with all Msi proteins containing two tandem RNA Recognition Motifs (RRMs) at their N-terminus [138, 140]. In mammals, there are two paralogous genes, *Msi1* and *Msi2*. *Msi1*, the most extensively studied member of the family, is preferentially expressed in stem and progenitor cells of various tissues such as brain, intestine, breast, hair follicle and pancreas [141–145]. In the brain, Msi1 is required for neural stem cell differentiation and multipotency [146]. Kawase *et al.* have discovered that the "D5E2" intronic enhancer located within intron six of *Msi1* locus regulates its expression specifically in neural stem/progenitor cells [147]. Functionally, a Msi1 promoter-driven GFP reporter can be used to isolate neural stem cells from human ventricular zone [148]. Almost all *Msi1*-GFP$^+$ cells expressed neural specific markers and are able to self-renew and generate both neuronal and glial cells, demonstrating that Msi1 marks human neural stem cells. Consistent with its expression pattern, *Msi1* is functionally essential for neural development in mice. *Msi1* gene knockout mice exhibit hydrocephaly due to a defect in ependymal cells but proliferation of neural progenitor cells surprisingly is not affected. Combined loss of *Msi1* and *Msi2*, however, leads to significant reduction in progenitor proliferation and impaired neuronal differentiation, suggesting functional redundancy between the two paralogous proteins in neural development. Extensive biochemical studies have elucidated the mechanism by which Msi1 controls stem cell properties in neural cells. Msi1 protein directly binds to the 3′-UTR of mRNAs encoding the Notch signaling inhibitor Numb and represses the

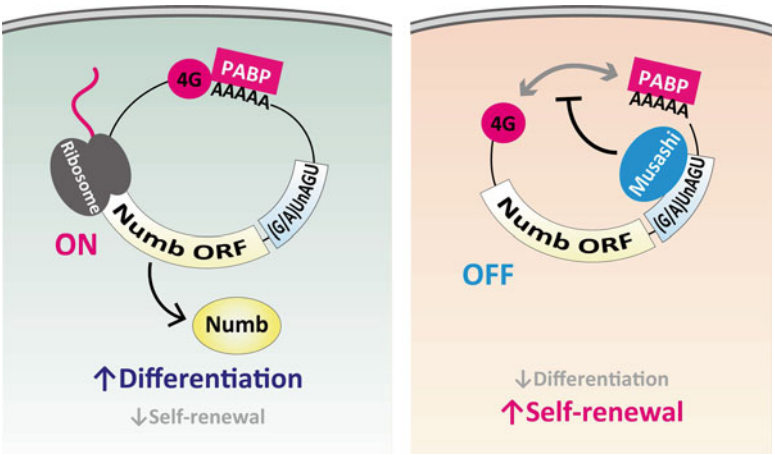


Fig. 7.6 Musashi-mediated translational repression. Musashi (MSI) negatively regulates mRNA translation by binding to the sequence $(G/A)U_{1-3}AGU$ within the 3′-UTRs of its target RNAs. The MSI binding to the transcript interferes with the interaction of the Poly(A) Binding Protein (PABP) with the translation initiation factor eIF4G, which leads to translational repression. By inhibiting expression of Numb which specifies differentiated cell fate, MSI maintains undifferentiated stem cell state

translation, leading to activated Notch signaling which in turn promotes self-renewal [149] (Fig. 7.6). More specifically, Msi1 represses 5′-cap-dependent translation of its target mRNAs by competing with eIF4G for PABP binding [150]. In neural progenitors, Msi1 also represses the translation of the cyclin-dependent kinase inhibitor p21/Waf1 via mRNA-binding, which promotes cell cycle progression [151]. Several studies have provided evidence for Msi1's role in stem cell self-renewal. In mammary progenitor cells, Msi1 stimulates the secretion of Proliferin-1, a growth stimulatory factor, which activates Wnt and Notch pathways to expand the progenitor cells [152]. In the intestine, Msi1 can mark crypt base columnar cells which are the intestinal epithelial stem cells expressing *Lgr5* [24, 28, 142, 153]. Overexpression of Msi1 in intestinal epithelial cells promotes progenitor proliferation through Wnt pathway activation induced by Msi1-mediated upregulation of Frat1, a potent activator of the canonical Wnt pathway [154].

Because Msi1 can activate the Wnt pathway in intestinal cells, it is not surprising that *Msi1* overexpression can confer tumorigenic potential on normal cells; i.e. *Msi1* acts as a potent oncogene. Indeed several studies have reported MSI1 overexpression in multiple tumor types including brain, mammary, endometrial and intestinal tumors, and MSI1 overexpression seems contributing to the stem cell population in these malignancies [142, 155–158]. For instance in glioblastoma, shRNA-mediated *MSI1* knockdown resulted in cell cycle arrest, impaired sphere formation and significantly inhibited in vivo tumor growth after xenograft [156]. Because *MSI1* knockdown leads to attenuated Notch and PI_3 kinase signaling, these two pathways may mediate MSI1-driven tumor development by promoting an

aberrant self-renewal program. In medulloblastoma, high MSI1 level predicts poor prognosis in patients and loss of *MSI1* induces tumor regression in vivo [159]. A recent study has revealed a novel regulatory link between Msi1 and the 26S proteasome in breast cancer and glioma stem cells [160]. The mRNA for NF-YA, a transcription factor that regulates expression of 26S proteasomal subunits, contains >20 copies of the putative Msi1 binding elements in the 3′-UTR. RIP assays confirm the direct binding of Msi1 protein to the NF-YA mRNA, which results in the reduced NF-YA expression and activity. In SUM159PT breast cancer cells, Msi1 expression is higher in the CSC population than in non-CSCs while the expression levels of NF-YA and the proteasomal subunits are significantly lower in CSCs. Accordingly, silencing *Msi1* in CSCs leads to increased proteasomal activity and loss of sphere-forming activity. Because resistance to proteasomal inhibitors such as bortezomib is prevalent in several tumors and likely caused by low 26S proteosomal activity, the Msi1-NF-YA axis might make a novel therapeutic target [161, 162].

Interestingly, through comprehensive small molecule screening, Ryder and colleagues have found that fatty acids inhibit the RNA-binding activity of MSI1 protein [163]. The 18–22 carbon omega-9 monounsaturated fatty acids such as oleic acid interact with RRM1 of MSI1, suggesting that the RNA-binding domain acts as a metabolite sensor. The *Scd1* mRNA for stearoyl-CoA desaturase that catalyzes omega-9 desaturation contain putative MSI1 binding elements, and MSI1 expression increased Scd1 protein expression, suggesting that Msi1 acts as a metabolite sensor and regulates *Scd1* translation dependent on cellular needs for unsaturated fatty acids. Collectively, the study identified a novel role for an RNA-binding protein in the regulation of gene expression by a cellular metabolite. Although the physiological significance of this regulation remains elusive, it is tempting to speculate that RBPs may modulate stem cell properties by sensing cellular changes of metabolic states.

The expression pattern of *Msi2*, the second Musashi paralogue in mammals, overlaps widely with that of *Msi1* [164]. In the hematopoietic tissue, however, *Msi2* is the sole member of the family expressed and is a functionally important regulator of differentiation in both normal and malignant hematopoietic cells [165, 166]. Within the hematopoietic system, Msi2 is highly expressed in the most primitive compartment, HSCs and MPPs. In contrast, very low expression is detected in lineage-committed progenitors and functioning mature cells of blood lineage except the burst-forming unit-erythroid (BFU-E) progenitors. *Msi2*-deficient mice are viable but born at reduced frequency and are smaller in size than their control littermates [166–168]. They show significant reduction in the numbers of HSCs, MPPs and leukocytes, which is exacerbated with age. The bone marrow repopulating ability of *Msi2* mutant HSCs is significantly impaired in the competitive reconstitution assay, indicating that stem cell renewal is affected in the absence of *Msi2*. Gene expression analysis of *Msi2*-deficient HSCs revealed that Msi2 maintains the stem cell compartment by regulating genes involved in cell proliferation, mitosis and progenitor differentiation [167, 168].

Msi2 has also been shown to play a significant role in regulating the pluripotency and self-renewal of ESCs [169]. ESCs express two isoforms of *Msi2*; the longer

canonical isoform and a shorter splice variant. Lentivirus-mediated downregulation of both isoforms promoted ESC differentiation and suppressed self-renewal. In contrast, overexpression of either isoform was not sufficient to rescue this effect, suggesting that both isoforms are required in regulation of ESC pluripotency. Because Msi2 has been shown to physically interact with Sox2, one of the critical pluripotency transcription factors, it is likely that Msi2 and Sox2 cooperate to suppress differentiation programs and promote ESC self-renewal [170].

In hematologic malignancies, *Msi2* expression levels are significantly elevated in the blast crisis phase of chronic myelogenouse leukemia (BC-CML) and *de novo* acute myeloid leukemia (AML) [166, 171, 172] (Fig. 7.6). NUP98-HOXA9, a fusion oncogene associated with both myeloid leukemias, can trigger *Msi2* expression, which in turn represses Numb protein translation. Genetic loss of *Msi2* in BC-CML increased Numb protein levels and, as a consequence, BC-CML cells became more differentiated and lost self-renewing activity in vivo. Importantly, *Msi2* is not only highly upregulated during human CML progression but is also a marker for poorer prognosis in human BC-CML. In *de novo* AMLs, MSI2 protein is found both in nucleus and cytoplasm, and high nuclear MSI2 levels predict unfavorable outcomes with statistical significance [171]. While Musashi proteins are known to act as translational regulators in the cytoplasm, their nuclear function remains to be elucidated in leukemogenesis and other malignancies.

3.5 HuR/Elav Family

The human antigen R (HuR) is a member of the embryonic lethal abnormal vision (Elav) family of RNA binding proteins and was first identified in *Drosophila* as an essential regulator of neural development [173, 174]. Proteins in the Elav family are highly conserved across multiple species from *C. elegans* to *H. sapiens*, with HuR being a human ortholog of *Drosophila* Elav and it thus has the alias *ELAVL1* (*Elav-like 1*). HuR protein contains three RRM motifs that are essential for binding AU-rich elements (AUUUA; ARE) found in the 3′-UTRs of various mRNAs. AREs have been shown to facilitate RNAs for rapid degradation [175]. By translocating from the nucleus to the cytoplasm, HuR directly binds to AREs in host mRNAs and stabilizes the transcript by inhibiting miRNA-mediated repression [176–178]. HuR can also negatively regulate protein expression of its targets by promoting miRNA-induced silencing complex (RISC)-mediated repression [179].

Whereas other Elav family proteins such as HuB, HuC and HuD are predominantly expressed in neuronal tissues, the HuR protein is ubiquitously expressed in intestine, spleen, thymus, colon and so on. Targeted gene disruption in mice revealed that homozygous null mutation in *Elavl1* results in no live-born animals, indicating its essential role in embryogenesis [180]. Postnatal deletion of *Elavl1* by ubiquitously expressed tamoxifen-inducible *Rosa26-CreER* leads to hematopoietic deficiencies and gastrointestinal tracts abnormalities. In the *Elavl1*-deficient mice, bone marrow, thymus, spleen and lymph nodes rapidly become hypocellular and circulating

leukocyte counts lower. The small intestine showed severe villus atrophy and disruption of the epithelial structure. Consistent with the profound hematopoietic phenotype, HuR is highly expressed in primitive hematopoitic progenitors. While HuR expression is low in quiescent HSCs, it is highly activated in MPPs especially within the nucleus. Loss of HuR expression induced p53 and its downstream target p21/Waf1 in both hematopoietic and intestinal cells, suggesting that HuR regulates p53-dependent cell proliferation and apoptotic response in the primitive progenitors. Increased sensitivity to radiation by the HuR deficiency is also consistent with its roles self-renewal since stem cell expansion is an essential step during regeneration after tissue damage [181–183].

HuR expression is perturbed in human cancers; overexpressed in breast, lung, ovarian, kidney, colon tumors and mesothelioma [184–188]. In breast cancer HuR is highly expressed in $CD44^{high}CD24^{low}$ cells with self-renewing activity, and cytoplasmic HuR protein levels, but not nuclear levels, significantly associated with higher tumor grades and larger tumor size [185, 186]. HuR protein physically interacts with β-catenin protein to stabilize mRNAs for hypoxia-inducible carbonic anhydrase 9 (*CA9*) and *SNAI2/SLUG* genes that are known to promote tumor cell survival and proliferation [189]. Upon HuR binding to β-catenin, the HuR-mRNA complex is recruited to the 40S ribosomal complex to facilitate translational activation of *CA9* and *SNAI2* mRNAs. HuR also enhances translation of β-catenin mRNA by direct binding [187, 190]. Taken together, the Wnt/β-catenin pathway seems to mediate the regulatory roles of HuR on CSC activity. Interestingly, HuR itself may regulate Wnt signals. HuR protein directly associates with the cyclooxygenase-2 (*COX2/PTGS2*) mRNA via an ARE in its 3′-UTR. Cyclooxygenase-2 is the enzyme responsible for the synthesis of prostanoids such as prostaglandin E2 (PGE2). PGE2 has been shown to positively regulate Wnt signaling in zebrafish and murine HSCs by stabilizing β-catenin via cAMP/PKA signaling [191, 192]. It remains to be elucidated whether HuR induction of cyclooxygenase-2 and PGE2 could contribute stem cell maintenance in vivo.

HuR also regulates the expression of other RBPs. In glioblastoma, HuR and Musashi expression levels are positively correlated; at the molecular level, HuR protein binds to *Msi1* mRNA at the 3′-UTR and positively regulates mRNA stability and translation [193]. Another HuR-regulated RBP is nucleolin (Ncl), a nucleolar RNA-binding protein with RRMs and functions in ribosomal RNA maturation and processing. RIP assays using anti-HuR antibody revealed HuR association with the 3′-UTR of *Ncl* mRNA, which then upregulates nucleolin protein expression by increasing translation [194]. In the same study, the authors identified *miR-494* targeting of Ncl mRNA at the 3′-UTR, suggesting that the miRNA and HuR compete with each other for modulation of nucleolin expression. In fact, recent genome-wide analysis has shown that 25 % of HuR-binding sites in the 3′-UTRs are shared with potential miRNA target sites, implying prevalent interplays between HuR and miRNAs [195]. Although there is little direct evidence supporting HuR function in self-renewal, HuR has been linked to stem cell regulatory proteins such as β-catenin and Musashi, with the studies presented here suggesting that HuR plays a role in normal and malignant stem cells.

3.6 FET Family

EWS, a founding member of the FET (FUS/TLS, EWS and TAF15/TAFII68) family of RNA-binding proteins, was first identified as the gene mutated in Ewing sarcoma [196, 197]. EWS protein contains an RRM and three Arg-Gly-Gly-rich stretches known as RGG motifs at its carboxyl terminal region. At the N-terminus, EWS contains the Ser-Tyr-Gln-Gly-rich (SYQG) motif that functions as a transcriptional activation domain. EWS protein shuttles between the nucleus and the cytoplasm and physically interacts with SF1 and U1C, components of the splicing machinery, suggesting its role in pre-mRNA processing and nuclear export of RNAs [198–201]. Several mRNA targets of EWS have been identified, including an Akt substrate PRAS40 [202]. By binding to the 3′-UTR of *PRAS40* mRNA, EWS negatively regulates *PRAS40* expression. Consistent with this finding, EWS overexpression results in reduced PRAS40 level, which then suppresses the proliferation and metastatic potential of Ewing sarcoma. Gene disruption studies have shown that loss of *Ews* leads to high perinatal lethality, arrested pre-B cell differentiation and meiotic defects during spermatogenesis [203]. In hematopoiesis, *Ews* is essential for stem cell quiescence and thus *Ews*-deficient mice exhibit HSC loss in bone marrow, impaired repopulating activity after transplantation and premature senescence in the blood compartment [204]. These studies have established an essential role of EWS in stem cell maintenance via regulation of cell cycle and quiescence.

Because of the history of discovery as described above, EWS and other FET family proteins are among the most extensively studied RNA-binding proteins in regard to oncogenesis. Ewing sarcoma is an aggressive bone tumor found in children and adolescents, characterized by a chromosomal translocation involving *EWSR1* gene encoding the EWS protein. Located on human chromosome 22, *EWSR1* is recurrently translocated to an ETS transcription factor gene, generating an oncogenic fusion protein between EWS and an ETS protein. *EWS-FLI1* is the most common fusion, occurring in >85 % of the reported cases of Ewing tumors [196]. The expression of *EWS-FLI1* alone is sufficient to initiate tumor development in both mouse and human bone-derived mesenchymal stem cells (MSCs), suggesting that the fusion protein is causative and that MSCs are the cells of origin for the tumor [205]. *EWS-FLI1* induces expression of polycomb group protein EZH2, self-renewal regulators OCT4, SOX2 and NANOG in human MSCs and initiates reprogramming of MSC into sarcoma cancer stem cells [205–207]. In a Cre-inducible transgenic mouse model, *EWS-FLI1* expression is shown to induce a rapid expansion of c-Kit$^+$ myeloid progenitors that can initiate leukemia after serial transplantation [208]. EWS fusion protein is thus capable of transforming normal cells by enhancing aberrant self-renewal.

Ewing sarcoma and myeloid leukemia are not the only malignancies in which FET family proteins become dysregulated. The *TLS/FUS* (Translocated in liposarcoma; Fused in sarcoma) gene was originally identified as a fusion partner of the *CHOP* gene in malignant liposarcoma with t(12;16) chromosomal translocation [209, 210]. In human myeloid leukemias with the recurrent t(16;21) translocation,

the *TLS/FUS* gene is translocated to the Ets-related gene *ERG*, generating a TLS-ERG chimeric protein. In both TLS-CHOP and TLS-ERG proteins, the RNA-binding domain at the C-terminus of TLS is replaced with the DNA binding domain from the corresponding transcription factors. In driving cellular transformation, the TLS fusion proteins act not only as chimeric transcription factors inducing deregulated gene transcription, but also affect RNA splicing. TLS protein recruits the SR family of splicing factors through interaction via its C-terminal domain whereas TLS-ERG fusion has lost its ability due to the lack of the interaction domain [211]. As a result, the TLS-ERG fusion inhibits pre-mRNA splicing mediated by the SR proteins and leads to alternate splicing variants of CD44. Like EWS-FLI1 fusion, ectopic expression of TLS-ERG also induces enhanced capacity for self-renewal and proliferation in human hematopoietic cells, implying its potent activity as an oncogene in diverse cell types [212].

Collectively these studies indicate that the FET RBPs and their fusion proteins hijack stem cell programs leading to cancer development. Of note, mutations in TLS/FUS genes have recently been linked to hereditary neurodegeneration in amyotrophic lateral sclerosis (ALS) and will be discussed in detail elsewhere in this series [213, 214].

3.7 Eukaryotic Translation Initiation Factor eIF4E

Unlike other RNA-binding proteins discussed so far, eukaryotic translation initiation factor eIF4E is unique in that it binds to RNAs via the 7-methyl-guanosine cap structure (m^7Gppp) at the 5′ ends of eukaryotic mRNAs. Together with other translation initiation factors such as eIF4G and eIF4A, eIF4E plays an essential role in recruiting the 43S ribosome complex during the initial step of translation in the cytoplasm, thereby promoting protein synthesis [215]. eIF4E is also found in the nucleus where it associates with and promotes the export of specific nuclear mRNAs independent of ongoing translation [216]. The eIF4E target mRNAs contain a structural element in their 3′-UTR known as 4E-SE (eIF4E sensitivity element), and many of the target mRNAs encode genes regulating cell cycle progression and survival such as *c-Myc*, *Mdm2* and *CyclinD1* [217]. In fact, eIF4E is capable of inducing malignant transformation in fibroblasts, and later studies have shown that eIF4E is highly expressed in multiple types of cancers and overexpression is often correlated with poor prognosis [218, 219]. In hematologic malignancies, where about 30 % of cases exhibit elevated eIF4E levels, the factor is almost exclusively localized in the nucleus with enhanced mRNA export activity [220, 221]. Ribavirin, a broad spectrum antiviral drug and a structural analog of the m^7G cap, impairs the eIF4E function both in the cytoplasm and the nucleus. Consequently this drug suppresses oncogenic transformation in vitro and is effective to induce favorable response in a group of patients with acute myeloid leukemia [220, 222]. In the blast crisis phase of CML, eIF4E overexpression activates the Wnt signaling pathway, which is essential for stem cell self-renewal in leukemia [223]. Together these studies suggest that inhibition of the RNA-binding protein eIF4E is a promising therapeutic strategy in human malignancies.

3.8 *PUF Family*

The PUF family of RBPs consists of evolutionarily conserved proteins found in unicellular and metazoan organisms [224, 225]. *Drosophila* Pumilio (Pum) and *C. elegans* FBF proteins are the first two members of this family and have been extensively studied [226, 227]. The PUF proteins are characterized by the presence of the RNA-binding domain termed Puf domain (previously known as Pumilio homology domain or PUM-HD), which contains eight imperfect Puf repeats consisting of 36 amino acids. The Puf domain recognizes 3′-UTRs containing a core sequence of 5′-UGUA-3′ in the target RNAs and inhibits gene expression by either facilitating degradation or interfering with the translation of target mRNA [228, 229].

The PUF proteins have multiple functions within the same species, plausibly by regulating different RNA targets. For instance, Pum regulates anterior-posterior axis formation and abdomen segmentation by repressing translation of *hunchback* (*hb*) in fruitflies [230]. This is achieved by Pum protein binding to specific sequences known as the Nanos Response Elements (NREs) in the 3′-UTR of *hb* mRNA and repressing its translation through recruitment of Nanos (Nos) and Brain Tumor (Brat) proteins by promoting deadenylation of the *hb* mRNA [231–234]. *Pum* also plays a significant role in asymmetric cell division in germline stem cells (GSCs) by allowing differential protein expression in each daughter cell. Fly embryos with *Pum* mutation have symmetric GSC division and generate two, instead of one, progenitor cells that differentiate into mature oocytes [235]. Interaction of Pum protein with multiple subunits of the CCR4-NOT mRNA deadenylation complex seems to be responsible for post-transcriptional gene regulation of its target mRNAs [236]. Mei-P26 is one of the direct targets of the Pum-Nos-CCR4 deadenylation complex in *Drosophila* GSCs [237]. *mei-P26* encodes a Trim-NHL domain-containing protein involved in GSC self-renewal and differentiation. Loss of *mei-P26* leads to the formation of tumors in cyst cells while overexpression renders loss of germ stem cells. The translational repression of *mei-P26* is mediated by direct binding of Pum to a degenerated motif (UGUAACAA) located in the 3′-UTR of *mei-P26*. The repression is dependent on interaction of Pum-Nos with the CCR4-NOT deadenylation complex, as is evident by the increased expression of Mei-P26 in *CCR4*-deficient mutants. This study demonstrates that GSC cell fate depends on Pum-Nos-CCR4 for precise regulation of Mei-P26 expression.

In humans as well as in mice two PUF family genes, *Pum1* and *Pum2*, have been identified. The two genes exhibit overall 75 % similarity and 91 % identity in their Puf RNA-binding domains [238]. Both mouse and human genes are expressed in multiple tissues and have largely overlapping expression patterns [239]. For instance, both are highly expressed in ESCs, fetal and adult ovary and testis. Within the reproductive organs, human PUM1 is ubiquitously expressed while PUM2 is enriched in primordial germ cells of fetal ovary and testis and early and late spermatocytes [240, 241]. Despite its specific expression pattern and ability to interact with Deleted-In-Azoospermia (DAZ) and DAZ-like (DAZL) proteins, which are required for maintenance and expansion of GSCs, *Pum2* is dispensable for reproduction and GSC maintenance in mice [241]. *Pum2* mutant mice have smaller testes but otherwise are

viable and fertile. On the other hand, loss of *Pum1*, which is ubiquitously expressed in the gonads, lead to impaired spermatogenesis with increased cell death in spermatocytes and infertility [242]. The increase in apoptosis is mediated by p53 activation and is partially rescued by *p53* loss in *Pum1* null background. Pum1 indirectly regulates p53 activity by directly binding to and repressing translation of mRNAs encoding p53 regulators such as Pias1 and Mdm2. These results indicate that *Pum1* functions in maintaining GSC homeostasis by modulating *p53* pathway in mammalian germline.

Other than germ cells, *Pum1* and *Pum2* are highly expressed in adult and fetal HSCs in mice [239]. Both genes are downregulated in early multipotent progenitors and become reactivated in mature blood cells, showing biphasic patterns of gene expression. This data suggests that *Pum1* and *Pum2* could be involved in HSC self-renewal as well as terminal differentiation of lineage restricted hematopoietic cells. However, more extensive studies are needed to determine the functional relevance of *Pum* genes in the blood lineage.

Pum2 seems to be an important regulator of other stem cells. It is expressed in human adipocyte-derived stem cells and MSCs from bone marrow [243]. Functionally, *Pum2* knockdown led to decreased cell proliferation with cell cycle arrest at the S phase, and *Pum2*-deficient cells showed differential expression of Pum2-associated mRNA targets involved in cell cycle and gene regulation. These findings imply that Pum2 is necessary for the maintenance and self-renewal of adipose-derived stem cells.

Several studies have shed new light on the roles of PUF family in stem cell function. Work on both *C. elegans* and human ESCs have identified a regulatory relationship between PUF proteins and MAPK expression [244]. Specifically, they show that PUM2 directly binds to NRE in the 3′-UTR of multiple MAPK mRNAs (MPK-1 in *C. elegans* and MAPK1/ERK2 and MAPK14/p38α in human) and represses their translation into functional proteins. Since ERK2 and p38 have been found to inhibit ESC self-renewal by promoting their differentiation, it is possible that PUM2 regulates ESCs via inhibition of MAPK protein expression [245].

4 Conclusion

Development of effective therapeutic strategies that enable regenerative medicine requires comprehensive understanding of molecular mechanisms governing stem cell maintenance and cell fate decision. To date developmental signals such as Wnt or pluripotency transcriptional factors like Oct4, Sox2 or Nanog have been identified as key regulators of stem cell self-renewal. In this chapter we have discussed functional roles played by RBPs in the regulation of stem cell homeostasis and their contribution to oncogenesis.

From a biological perspective, it is worth noting that the human genome encodes more than 1400 RBPs and at least one-third of them are expressed in embryonic or tissue stem cells (or both) and functionally required for their stem cell properties [246].

Their molecular functions in stem cells are diverse, including post-transcriptional gene regulation, small RNA biogenesis and non-coding RNA-mediated epigenetic regulation [130, 247–249]. These functions are mediated by the RBPs' ability to bind to multiple RNA targets either specifically or non-specifically. Although targets of most RBPs remain to be identified, technical advances in molecular genomics such as next-generation sequencing and single-cell analysis will soon allow us to decipher cellular regulatory network by RBPs and their target RNAs [250, 251].

From a clinical perspective, it is intriguing that multiple RBPs are expressed in both healthy and malignant stem cells and play significant roles in cell fate regulation [129, 165, 252]. The importance is twofold. First, by modulating RBP expression or function we may be able to find novel ways to keep stem cell behaviors under control, which is an essential component in utilizing stem cells or early progenitors in regenerative medicine. Second, the RBPs' ability to regulate cell fates is also critical during tumor development and progression because aberrant differentiation is one of major hallmarks of malignancy [166, 172, 253]. Consistently, expression levels of several RBPs in tumors predict unfavorable clinical outcomes, and more importantly, loss of function of these "oncogenic" RBPs are indeed effective in reversing cancer progression [82, 166, 254]. Therefore uncovering RBP functions and their RNA targets will definitely contribute to advance in our understanding of the mechanisms that underlie homeostasis, regeneration and oncogenesis.

Acknowledgment The authors would like to thank Nathaniel Beattie for helpful comments and suggestions on this manuscript. We apologize to those investigators whose works could not be cited due to space limitations.

References

1. Rossi DJ, Jamieson CHM, Weissman IL (2008) Stems cells and the pathways to aging and cancer. Cell 132:681–696
2. Takashima S, Hirose M, Ogonuki N, Ebisuya M, Inoue K, Kanatsu-Shinohara M, Tanaka T, Nishida E, Ogura A, Shinohara T (2013) Regulation of pluripotency in male germline stem cells by Dmrt1. Genes Dev 27:1949–1958
3. Yamanaka S (2012) Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 10:678–684
4. Osawa M, Hanada K-I, Hamada H, Nakauchi H (1996) Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science 273: 242–245
5. Morrison SJ, Weissman IL (1994) The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity 1:661–673
6. Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL (1997) Identification of a lineage of multipotent hematopoietic progenitors. Development 124:1929–1939
7. Till JE, McCulloch EA (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res 14:213–222
8. Uchida N, Weissman IL (1992) Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. J Exp Med 175:175–184

9. Adolfsson J, Borge OJ, Bryder D, Theilgaard-Mönch K, Astrand-Grundström I, Sitnicka E, Sasaki Y, Jacobsen S-E (2001) Upregulation of Flt3 expression within the bone marrow Lin(−)Sca1(+)c- kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. Immunity 15:659–669
10. Boyer SW, Schroeder AV, Smith-Berdan S, Forsberg EC (2011) All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. Cell Stem Cell 9:64–73
11. Kondo M, Weissman IL, Akashi K (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell 91:661–672
12. Orkin SH, Zon LI (2008) Hematopoiesis: an evolving paradigm for stem cell biology. Cell 132:631–644
13. Okada S, Nakauchi H, Nagayoshi K, Nishikawa S, Miura Y, Suda T (1992) In Vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. Blood 80:3044–3050
14. Chen C-Z, Li M, de Graaf D, Monti S, Göttgens B, Sanchez M-J, Lander ES, Golub TR, Green AR, Lodish HF (2002) Identification of endoglin as a functional marker that defines long-term repopulating hematopoietic stem cells. Proc Natl Acad Sci U S A 99:15468–15473
15. Christensen JL, Weissman IL (2001) Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc Natl Acad Sci U S A 98:14541–14546
16. Ema H, Sudo K, Seita J, Matsubara A, Morita Y, Osawa M, Takatsu K, Takaki S, Nakauchi H (2005) Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell 8:907–914
17. Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC (1996) Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med 183:1797–1806
18. Kiel MJ, Yilmaz ÖH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ (2005) SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell 121:1109–1121
19. Clevers H (2013) The intestinal crypt, a prototype stem cell compartment. Cell 154:274–284
20. Marshman E, Booth C, Potten CS (2002) The intestinal epithelial stem cell. Bioessays 24:91–98
21. Winton DJ, Ponder BA (1990) Stem-cell organization in mouse small intestine. Proc Biol Sci 241:13–18
22. Ireland H, Houghton C, Howard L, Winton DJ (2005) Cellular inheritance of a Cre-activated reporter gene to determine Paneth cell longevity in the murine small intestine. Dev Dyn 233:1332–1336
23. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, Danenberg E, Clarke AR, Sansom OJ, Clevers H (2009) Crypt stem cells as the cells-of-origin of intestinal cancer. Nature 457:608–611
24. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ et al (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449:1003–1007
25. Powell AE, Wang Y, Li Y, Poulin EJ, Means AL, Washington MK, Higginbotham JN, Juchheim A, Prasad N, Levy SE et al (2012) The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. Cell 149:146–158
26. Sangiorgi E, Capecchi MR (2008) Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet 40:915–920
27. Takeda N, Jain R, LeBoeuf MR, Wang Q, Lu MM, Epstein JA (2011) Interconversion between intestinal stem cell populations in distinct niches. Science 334:1420–1424
28. Muñoz J, Stange DE, Schepers AG, van de Wetering M, Koo B-K, Itzkovitz S, Volckmann R, Kung KS, Koster J, Radulescu S et al (2012) The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent "+4" cell markers. EMBO J 31:3079–3091

29. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ et al (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature 459:262–265
30. Yan KS, Chia LA, Li X, Ootani A, Su J, Lee JY, Su N, Luo Y, Heilshorn SC, Amieva MR (2012) The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. Proc Natl Acad Sci U S A 109:466–471
31. Dick JE (2008) Stem cell concepts renew cancer research. Blood 112:4793–4807
32. Jordan CT, Guzman ML, Noble M (2006) Cancer stem cells. N Engl J Med 355:1253–1261
33. Medema JP (2013) Cancer stem cells: the challenges ahead. Nat Cell Biol 15:338–344
34. Pierce GB, Dixon FJ (1959) Testicular teratomas. I. Demonstration of teratogenesis by metamorphosis of multipotential cells. Cancer 12:573–583
35. Bonnet D, Dick JE (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 3:730–737
36. Magee JA, Piskounova E, Morrison SJ (2012) Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell 21:283–296
37. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD, Dirks PB (2004) Identification of human brain tumour initiating cells. Nature 432:396–401
38. Tang B, Raviv A, Esposito D, Flanders KC, Daniel C, Nghiem BT, Garfield S, Lim L, Mannan P, Robles AI et al (2014) A flexible reporter system for direct observation and isolation of cancer stem cells. Stem Cell Reports 4(1):155–169
39. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA, Dick JE (1994) A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 367:645–648
40. Dalerba P, Dylla SJ, Park I-K, Liu R, Wang X, Cho RW, Hoey T, Gurney A, Huang EH, Simeone DM et al (2007) Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A 104:10158–10163
41. O'Brien CA, Pollett A, Gallinger S, Dick JE (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature 445:106–110
42. Ricci-Vitiani L, Lombardi DG, Pilozzi E, Biffoni M, Todaro M, Peschle C, De Maria R (2007) Identification and expansion of human colon-cancer-initiating cells. Nature 445:111–115
43. Jordan CT (2009) Cancer stem cells: controversial or just misunderstood? Cell Stem Cell 4:203–205
44. Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, Levine JE, Wang J, Hahn WC, Gilliland DG et al (2006) Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature 442:818–822
45. Wang Y, Krivtsov AV, Sinha AU, North TE, Goessling W, Feng Z, Zon LI, Armstrong SA (2010) The Wnt/beta-catenin pathway is required for the development of leukemia stem cells in AML. Science 327:1650–1653
46. Yeung J, Esposito MT, Gandillet A, Zeisig BB, Griessinger E, Bonnet D, So CWE (2010) β-Catenin mediates the establishment and drug resistance of MLL leukemic stem cells. Cancer Cell 18:606–618
47. Eberhart CG (2012) Three down and one to go: modeling medulloblastoma subgroups. Cancer Cell 21:137–138
48. Wang J, Wechsler-Reya RJ (2014) The role of stem cells and progenitors in the genesis of medulloblastoma. Exp Neurol 260:69–73
49. Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, Dewhirst MW, Bigner DD, Rich JN (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature 444:756–760
50. Corbin AS, Agarwal A, Loriaux M, Cortes J, Deininger MW, Druker BJ (2011) Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity. J Clin Invest 121:396–409
51. Zhang B, Strauss AC, Chu S, Li M, Ho Y, Shiang K-D, Snyder DS, Huettner CS, Shultz L, Holyoake T et al (2010) Effective targeting of quiescent chronic myelogenous leukemia stem

cells by histone deacetylase inhibitors in combination with imatinib mesylate. Cancer Cell 17:427–442
52. Reya T, Clevers H (2005) Wnt signalling in stem cells and cancer. Nature 434:843–850
53. Clevers H, Nusse R (2012) Wnt/β-catenin signaling and disease. Cell 149:1192–1205
54. Nusse R, Fuerer C, Ching W, Harnish K, Logan C, Zeng A, Berge DT, Kalani Y (2008) Wnt signaling and stem cell control. Cold Spring Harb Symp Quant Biol 73:59–66
55. Tolwinski NS, Wieschaus E (2004) Rethinking WNT signaling. Trends Genet 20:177–181
56. van Amerongen R, Mikels A, Nusse R (2008) Alternative wnt signaling is initiated by distinct receptors. Sci Signal 1:re9
57. Li VSW, Ng SS, Boersema PJ, Low TY, Karthaus WR, Gerlach JP, Mohammed S, Heck AJR, Maurice MM, Mahmoudi T et al (2012) Wnt signaling through inhibition of β-catenin degradation in an intact Axin1 complex. Cell 149:1245–1256
58. Liu C, Li Y, Semenov M, Han C, Baeg GH, Tan Y, Zhang Z, Lin X, He X (2002) Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. Cell 108:837–847
59. Korinek V, Barker N, Moerer P, van Donselaar E, Huls G, Peters PJ, Clevers H (1998) Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. Nat Genet 19:379–383
60. He XC, Zhang J, Tong W-G, Tawfik O, Ross J, Scoville DH, Tian Q, Zeng X, He X, Wiedemann LM et al (2004) BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt- beta-catenin signaling. Nat Genet 36:1117–1121
61. van Es JH, van Gijn ME, Riccio O, van den Born M, Vooijs M, Begthel H, Cozijnsen M, Robine S, Winton DJ, Radtke F et al (2005) Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. Nature 435:959–963
62. Bastide P, Darido C, Pannequin J, Kist R, Robine S, Marty-Double C, Bibeau F, Scherer G, Joubert D, Hollande F et al (2007) Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J Cell Biol 178:635–648
63. Blache P, van de Wetering M, Duluc I, Domon C, Berta P, Freund J-N, Clevers H, Jay P (2004) SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. J Cell Biol 166:37–47
64. Reya T, Duncan AW, Ailles L, Domen J, Scherer DC, Willert K, Hintz L, Nusse R, Weissman IL (2003) A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature 423:409–414
65. Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA (2008) Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev 22:746–755
66. Atlasi Y, Noori R, Gaspar C, Franken P, Sacchetti A, Rafati H, Mahmoudi T, Decraene C, Calin GA, Merrill BJ et al (2013) Wnt signaling regulates the lineage differentiation potential of mouse embryonic stem cells through Tcf3 down-regulation. PLoS Genet 9, e1003424
67. Bray SJ (2006) Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol 7:678–689
68. Honjo T (1996) The shortest path from the surface to the nucleus: RBP-Jκ/Su (H) transcription factor. Genes Cells 1:1–9
69. Ohsako S, Hyer J, Panganiban G, Oliver I, Caudy M (1994) Hairy function as a DNA-binding helix-loop-helix repressor of Drosophila sensory organ formation. Genes Dev 8:2743–2755
70. Vasyutina E, Lenhard DC, Wende H, Erdmann B, Epstein JA, Birchmeier C (2007) RBP-J (Rbpsuh) is essential to maintain muscle progenitor cells and to generate satellite cells. Proc Natl Acad Sci U S A 104:4443–4448
71. Bröhl D, Vasyutina E, Czajkowski MT, Griger J, Rassek C, Rahn H-P, Purfürst B, Wende H, Birchmeier C (2012) Colonization of the satellite cell niche by skeletal muscle progenitor cells depends on Notch signals. Dev Cell 23:469–481
72. Schroeder T, Fraser ST, Ogawa M, Nishikawa S, Oka C, Bornkamm GW, Nishikawa S-I, Honjo T, Just U (2003) Recombination signal sequence-binding protein Jkappa alters mesodermal cell fate decisions by suppressing cardiomyogenesis. Proc Natl Acad Sci U S A 100:4018–4023

73. Duncan AW, Rattis FM, DiMascio LN, Congdon KL, Pazianos G, Zhao C, Yoon K, Cook JM, Willert K, Gaiano N et al (2005) Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. Nat Immunol 6:314–322
74. Karanu FN, Murdoch B, Gallacher L, Wu DM, Koremoto M, Sakano S, Bhatia M (2000) The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. J Exp Med 192:1365–1372
75. Varnum-Finney B, Xu L, Brashem-Stein C, Nourigat C, Flowers D, Bakkour S, Pear WS, Bernstein ID (2000) Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. Nat Med 6:1278–1281
76. Lowell S, Benchoua A, Heavey B, Smith AG (2006) Notch promotes neural lineage entry by pluripotent embryonic stem cells. PLoS Biol 4, e121
77. Rock JR, Gao X, Xue Y, Randell SH, Kong Y-Y, Hogan BLM (2011) Notch-dependent differentiation of adult airway basal stem cells. Cell Stem Cell 8:639–648
78. Dreyfuss G, Kim VN, Kataoka N (2002) Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol 3:195–205
79. Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543
80. Han W, Xin Z, Zhao Z, Bao W, Lin X, Yin B, Zhao J, Yuan J, Qiang B, Peng X (2013) RNA-binding protein PCBP2 modulates glioma growth by regulating FHL3. J Clin Invest 123:2103–2118
81. Hu C-E, Tominaga K, Eiring AM, Liu Y-C, Srikantan S, Harb JG, Zhang H-D, Lee EK, Neviani P, Huang G-J et al (2014) The RNA-binding protein PCBP2 facilitates gastric carcinoma growth by targeting miR-34a. Biochem Biophys Res Commun 448:437–442
82. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, Iervolino A, Condorelli F, Gambacorti-Passerini C, Caligiuri MA et al (2002) BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet 30:48–58
83. Roychoudhury P, Paul RR, Chowdhury R, Chaudhuri K (2007) HnRNP E2 is downregulated in human oral cancer cells and the overexpression of hnRNP E2 induces apoptosis. Mol Carcinog 46:198–207
84. Melo JV, Barnes DJ (2007) Chronic myeloid leukaemia as a model of disease evolution in human cancer. Nat Rev Cancer 7:441–453
85. Chang JS, Santhanam R, Trotta R, Neviani P, Eiring AM, Briercheck E, Ronchetti M, Roy DC, Calabretta B, Caligiuri MA et al (2007) High levels of the BCR/ABL oncoprotein are required for the MAPK-hnRNP-E2 dependent suppression of C/EBP -driven myeloid differentiation. Blood 110:994–1003
86. Eiring AM, Harb JG, Neviani P, Garton C, Oaks JJ, Spizzo R, Liu S, Schwind S, Santhanam R, Hickey CJ et al (2010) miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. Cell 140:652–665
87. Yaniv K, Yisraeli JK (2002) The involvement of a conserved family of RNA-binding proteins in embryonic development and carcinogenesis. Gene 287:49–54
88. Bell JL, Wächter K, Mühleck B, Pazaitis N, Köhn M, Lederer M, Hüttelmaier S (2013) Insulin-like growth factor 2 mRNA-binding proteins (IGF2BPs): post-transcriptional drivers of cancer progression? Cell Mol Life Sci 70:2657–2675
89. Doyle GA, Betz NA, Leeds PF, Fleisig AJ, Prokipcak RD, Ross J (1998) The c-myc coding region determinant-binding protein: a member of a family of KH domain RNA-binding proteins. Nucleic Acids Res 26:5036–5044
90. Leeds P, Kren BT, Boylan JM, Betz NA, Steer CJ, Gruppuso PA, Ross J (1997) Developmental regulation of CRD-BP, an RNA-binding protein that stabilizes c-myc mRNA in vitro. Oncogene 14:1279–1286
91. Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC (1999) A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development. Mol Cell Biol 19:1262–1270
92. Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH (1997) Characterization of a beta-actin mRNA zipcode-binding protein. Mol Cell Biol 17:2158–2165

93. Hansen TV, Hammer NA, Nielsen J, Madsen M, Dalbaeck C, Wewer UM, Christiansen J, Nielsen FC (2004) Dwarfism and impaired gut development in insulin-like growth factor II mRNA- binding protein 1-deficient mice. Mol Cell Biol 24:4448–4464
94. Nielsen J, Adolph SK, Rajpert-De Meyts E, Lykke-Andersen J, Koch G, Christiansen J, Nielsen FC (2003) Nuclear transit of human zipcode-binding protein IMP1. Biochem J 376:383–391
95. Dimitriadis E, Trangas T, Milatos S, Foukas PG, Gioulbasanis I, Courtis N, Nielsen FC, Pandis N, Dafni U, Bardi G et al (2007) Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer. Int J Cancer 121:486–494
96. Elcheva I, Tarapore RS, Bhatia N, Spiegelman VS (2008) Overexpression of mRNA-binding protein CRD-BP in malignant melanomas. Oncogene 27:5069–5074
97. Hammer NA, Hansen TVO, Byskov AG, Rajpert-De Meyts E, Grøndahl ML, Bredkjaer HE, Wewer UM, Christiansen J, Nielsen FC (2005) Expression of IGF-II mRNA-binding proteins (IMPs) in gonads and testicular cancer. Reproduction 130:203–212
98. Kato T, Hayama S, Yamabuki T, Ishikawa N, Miyamoto M, Ito T, Tsuchiya E, Kondo S, Nakamura Y, Daigo Y (2007) Increased expression of insulin-like growth factor-II messenger RNA-binding protein 1 is associated with tumor progression in patients with lung cancer. Clin Cancer Res 13:434–442
99. Ross J, Lemm I, Berberet B (2001) Overexpression of an mRNA-binding protein in human colorectal cancer. Oncogene 20:6544–6550
100. Tessier CR, Doyle GA, Clark BA, Pitot HC, Ross J (2004) Mammary tumor induction in transgenic mice expressing an RNA-binding protein. Cancer Res 64:209–214
101. Ioannidis P, Mahaira LG, Perez SA, Gritzapis AD, Sotiropoulou PA, Kavalakis GJ, Antsaklis AI, Baxevanis CN, Papamichail M (2005) CRD-BP/IMP1 expression characterizes cord blood CD34+ stem cells and affects c-myc and IGF-II expression in MCF-7 cancer cells. J Biol Chem 280:20086–20093
102. Hamilton KE, Noubissi FK, Katti PS, Hahn CM, Davey SR, Lundsmith ET, Klein-Szanto AJ, Rhim AD, Spiegelman VS, Rustgi AK (2013) IMP1 promotes tumor growth, dissemination and a tumor-initiating cell phenotype in colorectal cancer cell xenografts. Carcinogenesis 34:2647–2654
103. Jeong W-J, Yoon J, Park J-C, Lee S-H, Lee S-H, Kaduwal S, Kim H, Yoon J-B, Choi K-Y (2012) Ras stabilization through aberrant activation of Wnt/β-catenin signaling promotes intestinal tumorigenesis. Sci Signal 5:ra30
104. Mongroo PS, Noubissi FK, Cuatrecasas M, Kalabis J, King CE, Johnstone CN, Bowser MJ, Castells A, Spiegelman VS, Rustgi AK (2011) IMP-1 displays cross-talk with K-Ras and modulates colon cancer cell survival through the novel proapoptotic protein CYFIP2. Cancer Res 71:2172–2182
105. Elcheva I, Goswami S, Noubissi FK, Spiegelman VS (2009) CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. Mol Cell 35:240–246
106. Noubissi FK, Elcheva I, Bhatia N, Shakoori A, Ougolkov A, Liu J, Minamoto T, Ross J, Fuchs SY, Spiegelman VS (2006) CRD-BP mediates stabilization of betaTrCP1 and c-myc mRNA in response to beta-catenin signalling. Nature 441:898–901
107. Boudoukha S, Cuvellier S, Polesskaya A (2010) Role of the RNA-binding protein IMP-2 in muscle cell motility. Mol Cell Biol 30:5710–5725
108. Brants JR, Ayoubi TAY, Chada K, Marchal K, Van de Ven WJM, Petit MMR (2004) Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. FEBS Lett 569:277–283
109. Fujii Y, Kishi Y, Gotoh Y (2013) IMP2 regulates differentiation potentials of mouse neocortical neural precursor cells. Genes Cells 18:79–89
110. Janiszewska M, Suvà ML, Riggi N, Houtkooper RH, Auwerx J, Clément-Schatlo V, Radovanovic I, Rheinbay E, Provero P, Stamenkovic I (2012) Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. Genes Dev 26:1926–1944

111. Liu W, Li Z, Xu W, Wang Q, Yang S (2013) Humoral autoimmune response to IGF2 mRNA-binding protein (IMP2/p62) and its tissue-specific expression in colon cancer. Scand J Immunol 77:255–260
112. Lederer M, Bley N, Schleifer C, Hüttelmaier S (2014) The role of the oncofetal IGF2 mRNA- binding protein 3 (IGF2BP3) in cancer. Semin Cancer Biol 29:3–12
113. Köbel M, Xu H, Bourne PA, Spaulding BO, Shih I-M, Mao T-L, Soslow RA, Ewanowich CA, Kalloger SE, Mehl E et al (2009) IGF2BP3 (IMP3) expression is a marker of unfavorable prognosis in ovarian carcinoma of clear cell subtype. Mod Pathol 22:469–475
114. Lin L, Zhang J, Wang Y, Ju W, Ma Y, Li L, Chen L (2013) Insulin-like growth factor-II mRNA- binding protein 3 predicts a poor prognosis for colorectal adenocarcinoma. Oncol Lett 6:740–744
115. Schaeffer DF, Owen DR, Lim HJ, Buczkowski AK, Chung SW, Scudamore CH, Huntsman DG, Ng SSW, Owen DA (2010) Insulin-like growth factor 2 mRNA binding protein 3 (IGF2BP3) overexpression in pancreatic ductal adenocarcinoma correlates with poor survival. BMC Cancer 10:59
116. Chen C-L, Tsukamoto H, Liu J-C, Kashiwabara C, Feldman D, Sher L, Dooley S, French SW, Mishra L, Petrovic L et al (2013) Reciprocal regulation by TLR4 and TGF-β in tumor-initiating stem- like cells. J Clin Invest 123:2832–2849
117. Liao B, Hu Y, Herrick DJ, Brewer G (2005) The RNA-binding protein IMP-3 is a translational activator of insulin-like growth factor II leader-3 mRNA during proliferation of human K562 leukemia cells. J Biol Chem 280:18517–18524
118. Song K, Wang H, Krebs TL, Danielpour D (2006) Novel roles of Akt and mTOR in suppressing TGF-beta/ALK5-mediated Smad3 activation. EMBO J 25:58–69
119. Chen P, Wang S-J, Wang H-B, Ren P, Wang X-Q, Liu W-G, Gu W-L, Li D-Q, Zhang T-G, Zhou C-J (2012) The distribution of IGF2 and IMP3 in osteosarcoma and its relationship with angiogenesis. J Mol Histol 43:63–70
120. Ueki A, Shimizu T, Masuda K, Yamaguchi SI, Ishikawa T, Sugihara E, Onishi N, Kuninaka S, Miyoshi K, Muto A (2012) Up-regulation of Imp3 confers in vivo tumorigenicity on murine osteosarcoma cells. PLoS One 7, e50621
121. Moss EG, Lee RC, Ambros V (1997) The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell 88:637–646
122. Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. Cell 147:1080–1091
123. Thornton JE, Gregory RI (2012) How does Lin28 let-7 control development and disease? Trends Cell Biol 22:474–482
124. Heo I, Lunde BM, Uren PJ, Yu F, Aoi T, Joo C, Moore C, Burns SC, Yao H, Yae K et al (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell 32:276–284
125. Hagan JP, Piskounova E, Gregory RI (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol 16:1021–1025
126. Newman MA, Thomson JM, Hammond SM (2008) Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. RNA 14:1539–1549
127. López de Silanes I, Yu J, Yu J, Fan J, Vodyanik MA, Vodyanik MA, Yang X, Smuga-Otto K, Smuga-Otto K, Zonderman AB et al (2007) Induced pluripotent stem cell lines derived from human somatic cells. Science 318:1917–1920
128. Moss EG, Tang L (2003) Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. Dev Biol 258:432–442
129. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J et al (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134:521–533
130. Melton C, Judson RL, Blelloch R (2010) Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature 463:621–626
131. Shyh-Chang N, Daley GQ (2013) Lin28: primal regulator of growth and metabolism in stem cells. Cell Stem Cell 12:395–406

132. Viswanathan SR, Powers JT, Einhorn W, Hoshida Y, Ng TL, Toffanin S, O'Sullivan M, Lu J, Phillips LA, Lockhart VL et al (2009) Lin28 promotes transformation and is associated with advanced human malignancies. Nat Genet 41:843–848
133. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J et al (2007) let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell 131:1109–1123
134. Cai W-Y, Wei T-Z, Luo Q-C, Wu Q-W, Liu Q-F, Yang M, Ye G-D, Wu J-F, Chen Y-Y, Sun G-B et al (2013) The Wnt-β-catenin pathway represses let-7 microRNA expression through transactivation of Lin28 to augment breast cancer stem cell expansion. J Cell Sci 126:2877–2889
135. Wu K, Jiao X, Li Z, Katiyar S, Casimiro MC, Yang W, Zhang Q, Willmarth NE, Chepelev I, Crosariol M et al (2011) Cell fate determination factor Dachshund reprograms breast cancer stem cell function. J Biol Chem 286:2132–2142
136. Dangi-Garimella S, Yun J, Eves EM, Newman M, Erkeland SJ, Hammond SM, Minn AJ, Rosner MR (2009) Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. EMBO J 28:347–358
137. Copley MR, Babovic S, Benz C, Knapp DJHF, Beer PA, Kent DG, Wohrer S, Treloar DQ, Day C, Rowe K et al (2013) The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat Cell Biol 15:916–925
138. Okano H, Imai T, Okabe M (2002) Musashi: a translational regulator of cell fate. J Cell Sci 115:1355–1359
139. Okano H, Kawahara H, Toriya M, Nakao K, Shibata S, Imai T (2005) Function of RNA-binding protein Musashi-1 in stem cells. Exp Cell Res 306:349–356
140. Sutherland JM, McLaughlin EA, Hime GR, Siddall NA (2013) The Musashi family of RNA-binding proteins: master regulators of multiple stem cell populations. Adv Exp Med Biol 786:233–245
141. Clarke RB, Spence K, Anderson E, Howell A, Okano H, Potten CS (2005) A putative human breast stem cell population is enriched for steroid receptor-positive cells. Dev Biol 277:443–456
142. Potten CS, Booth C, Tudor GL, Booth D, Brady G, Hurley P, Ashton G, Clarke R, Sakakibara S-I, Okano H (2003) Identification of a putative intestinal stem cell and early lineage marker; musashi-1. Differentiation 71:28–41
143. Sakakibara S, Imai T, Hamaguchi K, Okabe M, Aruga J, Nakajima K, Yasutomi D, Nagata T, Kurihara Y, Uesugi S et al (1996) Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. Dev Biol 176:230–242
144. Sugiyama-Nakagiri Y, Akiyama M, Shibata S, Okano H, Shimizu H (2006) Expression of RNA- binding protein Musashi in hair follicle development and hair cycle progression. Am J Pathol 168:80–92
145. Szabat M, Kalynyak TB, Lim GE, Chu KY, Yang YH, Asadi A, Gage BK, Ao Z, Warnock GL, Piret JM et al (2011) Musashi expression in β-cells coordinates insulin expression, apoptosis and proliferation in response to endoplasmic reticulum stress in diabetes. Cell Death Dis 2, e232
146. Sakakibara S-I, Nakamura Y, Yoshida T, Shibata S, Koike M, Takano H, Ueda S, Uchiyama Y, Noda T, Okano H (2002) RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. Proc Natl Acad Sci U S A 99:15194–15199
147. Kawase S, Imai T, Miyauchi-Hara C, Yaguchi K, Nishimoto Y, Fukami S-I, Matsuzaki Y, Miyawaki A, Itohara S, Okano H (2011) Identification of a novel intronic enhancer responsible for the transcriptional regulation of musashi1 in neural stem/progenitor cells. Mol Brain 4:14
148. Keyoung HM, Roy NS, Benraiss A, Louissaint A, Suzuki A, Hashimoto M, Rashbaum WK, Okano H, Goldman SA (2001) High-yield selection and extraction of two promoter-defined phenotypes of neural stem cells from the fetal human brain. Nat Biotechnol 19:843–850

149. Imai T, Tokunaga A, Yoshida T, Hashimoto M, Mikoshiba K, Weinmaster G, Nakafuku M, Okano H (2001) The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. Mol Cell Biol 21:3888–3900
150. Kawahara H, Imai T, Imataka H, Tsujimoto M, Matsumoto K, Okano H (2008) Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. J Cell Biol 181:639–653
151. Battelli C, Nikopoulos GN, Mitchell JG, Verdi JM (2006) The RNA-binding protein Musashi-1 regulates neural development through the translational repression of p21WAF-1. Mol Cell Neurosci 31:85–96
152. Wang XY, Yin Y, Yuan H, Sakamaki T, Okano H, Glazer RI (2008) Musashi1 modulates mammary progenitor cell expansion through proliferin-mediated activation of the Wnt and Notch pathways. Mol Cell Biol 28:3589–3599
153. Kayahara T, Sawada M, Takaishi S, Fukui H, Seno H, Fukuzawa H, Suzuki K, Hiai H, Kageyama R, Okano H (2003) Candidate markers for stem and early progenitor cells, Musashi-1 and Hes1, are expressed in crypt base columnar cells of mouse small intestine. FEBS Lett 535:131–135
154. Rezza A, Skah S, Roche C, Nadjar J, Samarut J, Plateroti M (2010) The overexpression of the putative gut stem cell marker Musashi-1 induces tumorigenesis through Wnt and Notch activation. J Cell Sci 123:3256–3265
155. Götte M, Wolf M, Staebler A, Buchweitz O, Kelsch R, Schüring AN, Kiesel L (2008) Increased expression of the adult stem cell marker Musashi-1 in endometriosis and endometrial carcinoma. J Pathol 215:317–329
156. Muto J, Imai T, Ogawa D, Nishimoto Y, Okada Y, Mabuchi Y, Kawase T, Iwanami A, Mischel PS, Saya H et al (2012) RNA-binding protein Musashi1 modulates glioma cell growth through the post-transcriptional regulation of Notch and PI3 kinase/Akt signaling pathways. PLoS One 7, e33431
157. Wang X-Y, Penalva LO, Yuan H, Linnoila RI, Lu J, Okano H, Glazer RI (2010) Musashi1 regulates breast tumor cell proliferation and is a prognostic indicator of poor survival. Mol Cancer 9:221
158. Li N, Yousefi M, Nakauka-Ddamba A, Li F, Vandivier L, Parada K, Woo D-H, Wang S, Naqvi AS, Rao S, Tobias J, Cedeno RJ, Minuesa G, Y K, Barlowe TS, Valvezan A, Shankar S, Deering RP, Klein PS, Jensen ST, Kharas MG, Gregory BD, Yu Z, Lengner CJ (2015) The Msi family of RNA-binding proteins function redundantly as intestinal oncoproteins. Cell Rep 13:2440–2455. doi:10.1016/j.celrep.2015.11.022
159. Vo DT, Subramaniam D, Remke M, Burton TL, Uren PJ, Gelfond JA, de Sousa Abreu R, Burns SC, Qiao M, Suresh U et al (2012) The RNA-binding protein Musashi1 affects medulloblastoma growth via a network of cancer-related genes and is an indicator of poor prognosis. Am J Pathol 181:1762–1772
160. Lagadec C, Vlashi E, Frohnen P, Alhiyari Y, Chan M, Pajonk F (2014) The RNA-binding protein Musashi-1 regulates proteasome subunit expression in breast cancer- and glioma-initiating cells. Stem Cells 32:135–144
161. Leung-Hagesteijn C, Erdmann N, Cheung G, Keats JJ, Stewart AK, Reece DE, Chung KC, Tiedemann RE (2013) Xbp1s-negative tumor B cells and pre-plasmablasts mediate therapeutic proteasome inhibitor resistance in multiple myeloma. Cancer Cell 24:289–304
162. Obeng EA, Carlson LM, Gutman DM, Harrington WJ, Lee KP, Boise LH (2006) Proteasome inhibitors induce a terminal unfolded protein response in multiple myeloma cells. Blood 107:4907–4916
163. Clingman CC, Deveau LM, Hay SA, Genga RM (2014) Allosteric inhibition of a stem cell RNA-binding protein by an intermediary metabolite. Elife 3
164. Sakakibara S, Nakamura Y, Satoh H, Okano H (2001) Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. J Neurosci 21:8091–8107

165. Hope KJ, Cellot S, Ting SB, Macrae T, Mayotte N, Iscove NN, Sauvageau G (2010) An RNAi screen identifies Msi2 and Prox1 as having opposite roles in the regulation of hematopoietic stem cell activity. Cell Stem Cell 7:101–113
166. Ito T, Kwon HY, Zimdahl B, Congdon KL, Blum J, Lento WE, Zhao C, Lagoo A, Gerrard G, Foroni L et al (2010) Regulation of myeloid leukaemia by the cell-fate determinant Musashi. Nature 466:765–768
167. de Andres-Aguayo L, Varas F, Kallin EM, Infante JF, Wurst W, Floss T, Graf T (2011) Musashi 2 is a regulator of the HSC compartment identified by a retroviral insertion screen and knockout mice. Blood 118:554–564
168. Park SM, Deering RP, Lu Y, Tivnan P, Lianoglou S, Al-Shahrour F, Ebert BL, Hacohen N, Leslie C, Daley GQ et al (2014) Musashi-2 controls cell fate, lineage bias, and TGF-β signaling in HSCs. J Exp Med 211:71–87
169. Wuebben EL, Mallanna SK, Cox JL, Rizzino A (2012) Musashi2 is required for the self-renewal and pluripotency of embryonic stem cells. PLoS One 7, e34827
170. Cox JL, Wilder PJ, Gilmore JM, Wuebben EL, Washburn MP, Rizzino A (2013) The SOX2-interactome in brain cancer cells identifies the requirement of MSI2 and USP9X for the growth of brain tumor cells. PLoS One 8, e62857
171. Byers RJ, Currie T, Tholouli E, Rodig SJ, Kutok JL (2011) MSI2 protein expression predicts unfavorable outcome in acute myeloid leukemia. Blood 118:2857–2867
172. Kharas MG, Lengner CJ, Al-Shahrour F, Bullinger L, Ball B, Zaidi S, Morgan K, Tam W, Paktinat M, Okabe R et al (2010) Musashi-2 regulates normal hematopoiesis and promotes aggressive myeloid leukemia. Nat Med 16:903–908
173. Campos AR, Rosen DR, Robinow SN, White K (1987) Molecular analysis of the locus elav in Drosophila melanogaster: a gene whose embryonic expression is neural specific. EMBO J 6:425
174. Colombrita C, Wilbert ML, Sutherland JM, Kim HH, Chang JS, Jordan CT, Wang J, Wu K, Zong F-Y, Erkinheimo T-L et al (2013) ELAV proteins along evolution: back to the nucleus? Mol Cell Neurosci 56:447–455
175. Barreau C, Paillard L, Osborne HB (2005) AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res 33:7138–7150
176. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell 125:1111–1124
177. Fan XC, Steitz JA (1998) HNS, a nuclear-cytoplasmic shuttling sequence in HuR. Proc Natl Acad Sci U S A 95:15293–15298
178. Keene JD (1999) Why is Hu where? Shuttling of early-response-gene messenger RNA subsets. Proc Natl Acad Sci U S A 96:5–7
179. Kim HH, Kuwano Y, Srikantan S, Lee EK, Martindale JL, Gorospe M (2009) HuR recruits let-7/RISC to repress c-Myc expression. Genes Dev 23:1743–1748
180. Ghosh M, Aguila HL, Michaud J, Ai Y, Wu M-T, Hemmes A, Ristimäki A, Guo C, Furneaux H, Hla T (2009) Essential role of the RNA-binding protein HuR in progenitor cell survival in mice. J Clin Invest 119:3530–3543
181. Abdelmohsen K, Pullmann R, Lal A, Kim HH, Galban S, Yang X, Blethrow JD, Walker M, Shubert J, Gillespie DA et al (2007) Phosphorylation of HuR by Chk2 regulates SIRT1 expression. Mol Cell 25:543–557
182. Kim HH, Abdelmohsen K, Lal A, Pullmann R, Yang X, Galban S, Srikantan S, Martindale JL, Blethrow J, Shokat KM et al (2008) Nuclear HuR accumulation through phosphorylation by Cdk1. Genes Dev 22:1804–1815
183. Mazan-Mamczarz K, Hagner PR, Zhang Y, Dai B, Lehrmann E, Becker KG, Keene JD, Gorospe M, Liu Z, Gartenhaus RB (2011) ATM regulates a DNA damage response posttranscriptional RNA operon in lymphocytes. Blood 117:2441–2450
184. Abdelmohsen K, Gorospe M (2010) Posttranscriptional regulation of cancer traits by HuR. Wiley Interdiscip Rev RNA 1:214–229

185. Denkert C, Weichert W, Winzer K-J, Müller B-M, Noske A, Niesporek S, Kristiansen G, Guski H, Dietel M, Hauptmann S (2004) Expression of the ELAV-like protein HuR is associated with higher tumor grade and increased cyclooxygenase-2 expression in human breast carcinoma. Clin Cancer Res 10:5580–5586
186. Erkinheimo T-L, Lassus H, Sivula A, Sengupta S, Furneaux H, Hla T, Haglund C, Butzow R, Ristimäki A (2003) Cytoplasmic HuR expression correlates with poor outcome and with cyclooxygenase 2 expression in serous ovarian carcinoma. Cancer Res 63:7591–7594
187. López de Silanes I, Fan J, Yang X, Zonderman AB, Potapova O, Pizer ES, Gorospe M (2003) Role of the RNA-binding protein HuR in colon carcinogenesis. Oncogene 22:7146–7154
188. Sengupta S, Jang B-C, Wu M-T, Paik J-H, Furneaux H, Hla T (2003) The RNA-binding protein HuR regulates the expression of cyclooxygenase-2. J Biol Chem 278:25227–25233
189. D'Uva G, Bertoni S, Lauriola M, De Carolis S, Pacilli A, D'Anello L, Santini D, Taffurelli M, Ceccarelli C, Yarden Y (2013) Beta-catenin/HuR post-transcriptional machinery governs cancer stem cell features in response to hypoxia. PLoS One 8, e80742
190. Chou S-D, Murshid A, Eguchi T, Gong J, Calderwood SK (2014) HSF1 regulation of β-catenin in mammary cancer cells through control of HuR/elavL1 expression. Oncogene 34(17):2178–2188
191. Goessling W, North TE, Loewer S, Lord AM, Lee S, Stoick-Cooper CL, Weidinger G, Puder M, Daley GQ, Moon RT et al (2009) Genetic interaction of PGE2 and Wnt signaling regulates developmental specification of stem cells and regeneration. Cell 136:1136–1147
192. North TE, Goessling W, Walkley CR, Lengerke C, Kopani KR, Lord AM, Weber GJ, Bowman TV, Jang I-H, Grosser T et al (2007) Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. Nature 447:1007–1011
193. Vo DT, Abdelmohsen K, Martindale JL, Qiao M, Tominaga K, Burton TL, Gelfond JAL, Brenner AJ, Patel V, Trageser D et al (2012) The oncogenic RNA-binding protein Musashi1 is regulated by HuR via mRNA translation and stability in glioblastoma cells. Mol Cancer Res 10:143–155
194. Tominaga K, Srikantan S, Lee EK, Subaran SS, Martindale JL, Abdelmohsen K, Gorospe M (2011) Competitive regulation of nucleolin expression by HuR and miR-494. Mol Cell Biol 31:4219–4231
195. Uren PJ, Burns SC, Ruan J, Singh KK, Smith AD, Penalva LOF (2011) Genomic analyses of the RNA-binding protein Hu antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. J Biol Chem 286:37063–37066
196. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, Rouleau G (1992) Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature 359:162–165
197. Janknecht R (2005) EWS-ETS oncoproteins: the linchpins of Ewing tumors. Gene 363:1–14
198. Knoop LL, Baker SJ (2000) The splicing factor U1C represses EWS/FLI-mediated transactivation. J Biol Chem 275:24865–24871
199. Knoop LL, Baker SJ (2001) EWS/FLI alters 5′-splice site selection. J Biol Chem 276:22317–22322
200. Zakaryan RP, Gehring H (2006) Identification and characterization of the nuclear localization/retention signal in the EWS proto-oncoprotein. J Mol Biol 363:27–38
201. Zhang D, Paley AJ, Childs G (1998) The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. J Biol Chem 273:18086–18091
202. Huang L, Nakai Y, Kuwahara I, Matsumoto K (2012) PRAS40 is a functionally critical target for EWS repression in Ewing sarcoma. Cancer Res 72:1260–1269
203. Li H, Watford W, Li C, Parmelee A, Bryant MA, Deng C, O'Shea J, Lee SB (2007) Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. J Clin Invest 117:1314–1323
204. Cho J, Shen H, Yu H, Li H, Cheng T, Lee SB, Lee BC (2011) Ewing sarcoma gene Ews regulates hematopoietic stem cell senescence. Blood 117:1156–1166

205. Riggi N, Suvà M-L, Suvà D, Cironi L, Provero P, Tercier S, Joseph J-M, Stehle J-C, Baumer K, Kindler V et al (2008) EWS-FLI-1 expression triggers a Ewing's sarcoma initiation program in primary human mesenchymal stem cells. Cancer Res 68:2176–2185
206. Riggi N, Suvà M-L, De Vito C, Provero P, Stehle J-C, Baumer K, Cironi L, Janiszewska M, Petricevic T, Suvà D et al (2010) EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. Genes Dev 24:916–932
207. Suvà M-L, Riggi N, Stehle J-C, Baumer K, Tercier S, Joseph J-M, Suvà D, Clément V, Provero P, Cironi L et al (2009) Identification of cancer stem cells in Ewing's sarcoma. Cancer Res 69:1776–1781
208. Torchia EC, Boyd K, Rehg JE, Qu C, Baker SJ (2007) EWS/FLI-1 induces rapid onset of myeloid/erythroid leukemia in mice. Mol Cell Biol 27:7918–7934
209. Crozat A, Aman P, Mandahl N, Ron D (1993) Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature 363:640–644
210. Rabbitts TH, Forster A, Larson R, Nathan P (1993) Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. Nat Genet 4:175–180
211. Yang L, Embree LJ, Hickstein DD (2000) TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. Mol Cell Biol 20:3345–3354
212. Pereira DS, Dorrell C, Ito CY, Gan OI, Murdoch B, Rao VN, Zou JP, Reddy ES, Dick JE (1998) Retroviral transduction of TLS-ERG initiates a leukemogenic program in normal human hematopoietic cells. Proc Natl Acad Sci U S A 95:8239–8244
213. Lagier-Tourenne C, Polymenidou M, Cleveland DW (2010) TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum Mol Genet 19:R46–R64
214. Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling S-C, Liang TY, Mazur C et al (2012) Divergent roles of ALS-linked proteins FUS/TLS and TDP- 43 intersect in processing long pre-mRNAs. Nat Neurosci 15:1488–1497
215. Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell 136:731–745
216. Culjkovic B, Topisirovic I, Skrabanek L, Ruiz-Gutierrez M, Borden KLB (2006) eIF4E is a central node of an RNA regulon that governs cellular proliferation. J Cell Biol 175:415–426
217. Culjkovic B, Topisirovic I, Skrabanek L, Ruiz-Gutierrez M, Borden KLB (2005) eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3′UTR. J Cell Biol 169:245–256
218. Lazaris-Karatzas A, Montine KS, Sonenberg N (1990) Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5′ cap. Nature 345:544–547
219. Mamane Y, Petroulakis E, Rong L, Yoshida K, Ler LW, Sonenberg N (2004) eIF4E- from translation to transformation. Oncogene 23:3172–3179
220. Assouline S, Culjkovic B, Cocolakis E, Rousseau C, Beslu N, Amri A, Caplan S, Leber B, Roy D-C, Miller WH et al (2009) Molecular targeting of the oncogene eIF4E in acute myeloid leukemia (AML): a proof-of-principle clinical trial with ribavirin. Blood 114:257–260
221. Topisirovic I, Guzman ML, McConnell MJ, Licht JD, Culjkovic B, Neering SJ, Jordan CT, Borden KLB (2003) Aberrant eukaryotic translation initiation factor 4E-dependent mRNA transport impedes hematopoietic differentiation and contributes to leukemogenesis. Mol Cell Biol 23:8992–9002
222. Kentsis A, Topisirovic I, Culjkovic B, Shao L, Borden KL (2004) Ribavirin suppresses eIF4E- mediated oncogenic transformation by physical mimicry of the 7-methyl guanosine mRNA cap. Proc Natl Acad Sci U S A 101:18105–18110
223. Lim S, Saw TY, Zhang M, Janes MR, Nacro K, Hill J, Lim AQ, Chang C-T, Fruman DA, Rizzieri DA et al (2013) Targeting of the MNK–eIF4E axis in blast crisis chronic myeloid leukemia inhibits leukemia stem cell function. Proc Natl Acad Sci U S A 110:E2298–E2307
224. Quenault T, Lithgow T, Traven A (2011) PUF proteins: repression, activation and mRNA localization. Trends Cell Biol 21:104–112

225. Wickens M, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3′ UTR regulation as a way of life. Trends Genet 18:150–157
226. Zamore PD, Williamson JR, Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA 3:1421–1433
227. Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP (1997) A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. Nature 390:477–484
228. Edwards TA, Pyle SE, Wharton RP, Aggarwal AK (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs. Cell 105:281–289
229. White EK, Moore-Jarrett T, Ruley HE (2001) PUM2, a novel murine puf protein, and its consensus RNA-binding site. RNA 7:1855–1866
230. Struhl G, Johnston P, Lawrence PA (1992) Control of Drosophila body pattern by the hunchback morphogen gradient. Cell 69:237–249
231. Loedige I, Stotz M, Qamar S, Kramer K, Hennig J, Schubert T, Löffler P, Längst G, Merkl R, Urlaub H et al (2014) The NHL domain of BRAT is an RNA-binding domain that directly contacts the hunchback mRNA for regulation. Genes Dev 28:749–764
232. Sonoda J, Wharton RP (1999) Recruitment of Nanos to hunchback mRNA by Pumilio. Genes Dev 13:2704–2712
233. Sonoda J, Wharton RP (2001) Drosophila Brain Tumor is a translational repressor. Genes Dev 15:762–773
234. Wreden C, Verrotti AC, Schisa JA, Lieberfarb ME, Strickland S (1997) Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. Development 124:3015–3023
235. Forbes A, Lehmann R (1998) Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. Development 125:679–690
236. Van Etten J, Schagat TL, Hrit J, Weidmann CA, Brumbaugh J, Coon JJ, Goldstrohm AC (2012) Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. J Biol Chem 287:36370–36383
237. Joly W, Chartier A, Rojas-Rios P, Busseau I, Simonelig M (2013) The CCR4 deadenylase acts with Nanos and Pumilio in the fine-tuning of Mei-P26 expression to promote germline stem cell self-renewal. Stem Cell Reports 1:411–424
238. Spassov DS, Jurecic R (2002) Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins. Gene 299:195–204
239. Spassov DS, Jurecic R (2003) Mouse Pum1 and Pum2 genes, members of the Pumilio family of RNA-binding proteins, show differential expression in fetal and adult hematopoietic stem cells and progenitors. Blood Cells Mol Dis 30:55–69
240. Moore FL, Jaruzelska J, Fox MS, Urano J, Firpo MT, Turek PJ, Dorfman DM, Pera RAR (2003) Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. Proc Natl Acad Sci U S A 100:538–543
241. Xu EY, Chang R, Salmon NA, Reijo Pera RA (2007) A gene trap mutation of a murine homolog of the Drosophila stem cell factor Pumilio results in smaller testes but does not affect litter size or fertility. Mol Reprod Dev 74:912–921
242. Chen D, Zheng W, Lin A, Uyhazi K, Zhao H, Lin H (2012) Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis. Curr Biol 22:420–425
243. Shigunov P, Sotelo-Silveira J, Kuligovski C, de Aguiar AM, Rebelatto CK, Moutinho JA, Brofman PS, Krieger MA, Goldenberg S, Munroe D et al (2012) PUMILIO-2 is involved in the positive regulation of cellular proliferation in human adipose-derived stem cells. Stem Cells Dev 21:217–227
244. Lee M-H, Hook B, Pan G, Kershner AM, Merritt C, Seydoux G, Thomson JA, Wickens M, Kimble J (2007) Conserved regulation of MAP kinase expression by PUF RNA-binding proteins. PLoS Genet 3, e233

245. Armstrong L, Hughes O, Yung S, Hyslop L, Stewart R, Wappler I, Peters H, Walter T, Stojkovic P, Evans J et al (2006) The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. Hum Mol Genet 15:1894–1913
246. Kwon SC, Yi H, Eichelbaum K, Föhr 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(9):1122–1130
247. Davidovich C, Zheng L, Goodrich KJ, Cech TR (2013) Promiscuous RNA-binding by Polycomb repressive complex 2. Nat Struct Mol Biol 20:1250–1257
248. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477:295–300
249. Wilbert ML, Huelga SC, Kapeli K, Stark TJ, Liang TY, Chen SX, Yan BY, Nathanson JL, Hutt KR, Lovci MT et al (2012) LIN28 Binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. Mol Cell 48:195–206
250. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149(6):1393–1406
251. Paz I, Kosti I, Ares M, Cline M, Mandel-Gutfreund Y (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Res 42(Web Server issue):W361–W367
252. Kechavarzi B, Janga SC (2014) Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome Biol 15:R14
253. Zhu Z, Khan MA, Weiler M, Blaes J, Jestaedt L, Geibert M, Zou P, Gronych J, Bernhardt O, Korshunov A et al (2014) Targeting self-renewal in high-grade brain tumors leads to loss of brain tumor stem cells and prolonged survival. Cell Stem Cell 15:185–198
254. Jonsson L, Bergman J, Nodin B, Manjer J, Pontén F, Uhlén M, Jirström K (2011) Low RBM3 protein expression correlates with tumour progression and poor prognosis in malignant melanoma: an analysis of 215 cases from the Malmö Diet and Cancer Study. J Transl Med 9:114

Chapter 8
Controlling the Editor: The Many Roles of RNA-Binding Proteins in Regulating A-to-I RNA Editing

Michael C. Washburn and Heather A. Hundley

Abstract RNA editing is a cellular process used to expand and diversify the RNA transcripts produced from a generally immutable genome. In animals, the most prevalent type of RNA editing is adenosine (A) to inosine (I) deamination catalyzed by the ADAR family. Throughout development, A-to-I editing levels increase while ADAR expression is constant, suggesting cellular mechanisms to regulate A-to-I editing exist. Furthermore, in several disease states, ADAR expression levels are similar to the normal state, but A-to-I editing levels are altered. Therefore, understanding how these enzymes are regulated in normal tissues and misregulated in disease states is of profound importance. This chapter will both discuss how to identify A-to-I editing sites across the transcriptome and explore the mechanisms that regulate ADAR editing activity, with particular focus on the diverse types of RNA-binding proteins implicated in regulating A-to-I editing *in vivo*.

Keywords RNA editing • ADARs • Inosine • RNA-binding proteins • Splicing

1 Introduction to RNA Editing

Despite all cells of the human body containing the same genetic material, there are many specialized cell types that perform the specific functions required to maintain homeostasis. Cells have evolved a variety of mechanisms to achieve molecular complexity, including transcriptional regulation, mRNA processing, and post-translational protein modifications. In regards to RNA processing, nearly 95 % of human genes are predicted to undergo alternative splicing; therefore, much attention has focused on how alternative splicing generates diverse transcripts [1].

M.C. Washburn
Department of Biology, Indiana University, Bloomington, IN 47405, USA

H.A. Hundley (✉)
Medical Sciences Program, Indiana University, Bloomington, IN 47405, USA
e-mail: hahundle@indiana.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_8

However, other mechanisms that change the information in the mRNA transcript, such as RNA editing, also play a major role in generating diverse transcriptomes from a static genome [2, 3].

RNA editing was discovered in the late 1980s, when specific RNA nucleotides present in cellular transcripts were found to differ in sequence from their genomic origin [4, 5]. These alterations included nucleotide insertions, deletions and base changes and are now generally referred to as RNA editing [6]. Base changes are the most widespread type of RNA editing and occur both in plant organelles and the nucleus of higher eukaryotes [7]. In humans, two types of base changes present within mRNA transcripts have been well studied: adenosine (A) to inosine (I) and cytidine (C) to uridine (U) [8, 9]. Both of these editing events involve hydrolytic deamination reactions that alter the base-pairing potential of the nucleotide. Recent transcriptome-wide studies have determined that adenosine deamination occurs in over two-thirds of human mRNAs, with the predicted number of A-to-I editing sites in the human transcriptome reaching 100 million [10]. In contrast, only approximately 100 C-to-U editing events have been identified in human transcripts [11]. Although C-to-U editing is known to play critical roles in regulating gene expression, see [12] for review, this chapter will focus on A-to-I RNA editing.

Catalyzed by the adenosine deaminase that act on RNA (ADAR) family, A-to-I editing occurs in regions of double-stranded RNA (dsRNA) [9, 13]. As inosine preferentially base-pairs with cytidine, A-to-I editing can have a variety of effects depending on the location of the editing event within the RNA (Fig. 8.1). A-to-I editing in codons can change the amino acid composition of the protein [14]. Most sequences critical for proper splicing contain adenosines, including the 5′ donor sequence, the 3′ acceptor sequence and the branch point adenosine; thus, A-to-I editing within introns and exons can alter splicing [15]. A-to-I editing events occurring in 3′ UTRs can modify small RNA (siRNA and miRNA) binding sites [16]. In addition, as miRNA and siRNA precursors form dsRNA structures, these molecules are also known targets for ADARs. Depending on the location of editing within the miRNA/siRNA precursors, editing can affect small RNA processing and/or alter binding of the small RNA to target mRNA [17–19].

A-to-I editing is essential for normal development, as mice lacking ADARs die in utero or shortly after birth [20, 21]. In addition, loss of ADARs results in severe neurological phenotypes, including seizures in mammals, neurodegeneration and uncoordinated movement in flies, and defective chemotaxis in worms [20, 22, 23]. Alterations in A-to-I editing have been linked to several neurological diseases including epilepsy, schizophrenia, amyotrophic lateral sclerosis, Alzheimer's disease and brain cancers [24–28]. However, expression levels of the ADAR enzymes do not directly correlate with levels of RNA editing observed in patients suffering from these diseases [24, 29, 30]. Similarly, A-to-I editing levels increase during normal development independent of ADAR protein expression levels [31], suggesting that there are cellular mechanisms that exist to regulate ADAR editing activity.

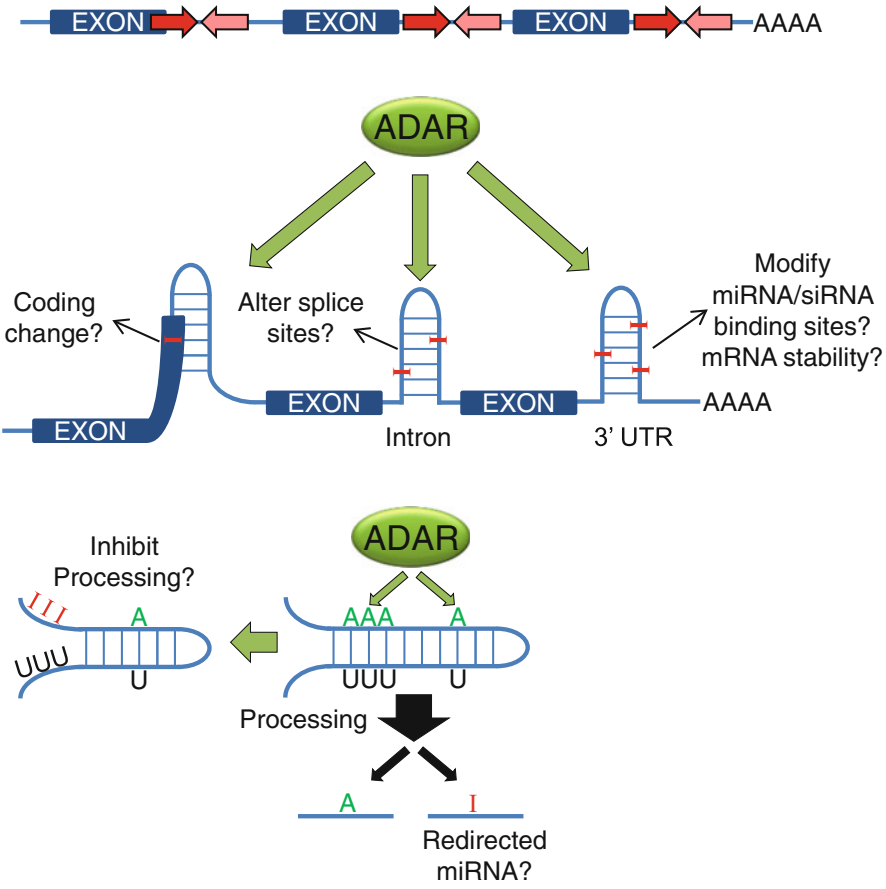


Fig. 8.1 Consequences and outcomes of A-to-I editing events in mRNAs and small RNAs. (**a**) ADAR substrates are comprised of complementary regions (represented by *arrows*) present in inverse orientation within in one mRNA, which can fold into a dsRNA hairpin following transcription. (**b**) ADARs can deaminate adenosine present in dsRNA structures within different regions of mRNAs. Within exonic sequences, editing events (represented by I) can alter the coding potential of the modified exon. Intronic editing events can generate alternative splicing products due to modification of 5′ donor sites, branch point adenosines, 3′ acceptor sites or by creating/removing splicing factor binding sites. A-to-I editing events in 3′ UTRs have been demonstrated to alter miRNA-binding sites. While the function of many editing events in 3′ UTRs have yet to be determined, these editing sites could also alter mRNA localization, export or stability by interfering with other RBP pathways. (**c**) Editing in miRNA precursors can generate two possible outcomes: (*left of green arrow*) altering the processing efficiency by Drosha/Dicer or (*below black arrow*) retargeting the miRNA to novel target mRNAs. The example shown is for miRNA editing, but the same outcomes apply to editing of siRNA precursors

1.1 Influence of ADAR Protein Domains on A-to-I Editing

All ADAR family members are composed of one or more N-terminal dsRNA-binding domains (dsRBDs) and a C-terminal deaminase domain [13]. ADARs are present in animals from cnidarians to humans, with one to four ADAR family members encoded in a given genome [32]. Even though ADAR family members share a high level of sequence homology, ADAR family members exhibit distinct, but partially overlapping preferences for target adenosines [33]. Several high-throughput studies suggest that ADARs prefer an adenosine, uridine or cytidine 5′ of the target adenosine while favoring guanosine 3′ to the target adenosine [10, 34, 35]. In addition, human ADARs prefer cytidine as the nucleotide opposing the target adenosine [36]. As many of these nucleotide preferences are also observed from *in vitro* editing reactions with ADARs lacking the dsRNA-binding domains, it is presumed that the deaminase domain prefers a specific sequence context surrounding the target adenosine [37]. Moreover, the 5′ nucleotide preferences, but not the 3′ preference, are shared between human ADAR1 and ADAR2 demonstrating their distinct, but overlapping preferences. These nucleotide preferences are hypothesized to result from their influence on flipping the adenosine out of dsRNA, a key step in the deamination reaction [38].

While many of these nucleotide preferences have been attributed to the deaminase domains, the dsRBDs also play a role in selecting target adenosines [39]. For example, a chimeric protein containing the deaminase domain of human ADAR1 and the dsRBDs from another dsRNA-binding protein, human PKR, cannot efficiently edit natural ADAR substrates [40]. In addition, the full length version of human ADAR2, but not the ADAR2 deaminase domain alone, demonstrates a preference for a guanosine 3′ to the target adenosine [37]. Further support for a role of the dsRBDs in selecting ADAR target adenosines comes from recent structural studies [41, 42]. Specifically, the dsRBDs of mammalian ADAR2 were shown to be capable of sequence specific contacts with the nucleotide 3′ to the target adenosine, perhaps explaining why ADAR2 has a preference for guanosine 3′ of the adenosine [42]. However, transcriptome-wide binding site identification experiments have not been published for ADARs in any organism, so it is unclear how critical the neighboring nucleotides are in selecting target adenosines *in vivo*.

2 Identification of RNA Editing Sites

To date, studies aimed at identifying RNA editing sites have uncovered millions of editing sites in organisms ranging from worms to humans [43]. The first endogenous ADAR editing targets were discovered by chance when researchers noticed discrepancies between sequences encoded by the genome versus cDNAs of cellular RNAs [44]. Since inosine is reverse transcribed as guanosine in cDNA, an A-to-G change between a gene and its cDNA can be indicative of A-to-I editing (Fig. 8.2).

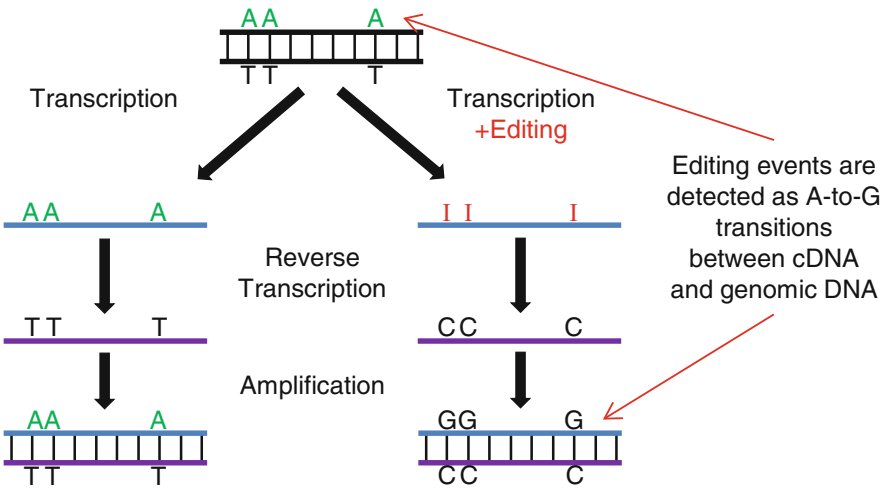


Fig. 8.2 How to detect A-to-I editing events in cellular RNAs. Transcription from genomic DNA (*black*, *Top*) yields mRNA (*blue*). If editing occurs (*right hand panel*), adenosine (A) is deaminated to inosine (I). Reverse transcriptase recognizes inosine as guanosine resulting in cytidine (C) in the cDNA (*purple*). Following PCR amplification, the A-to-I editing event will appear as an A-to-G transition when comparing the sequence of genomic DNA to the PCR amplified cDNA

However, if the cDNA sequence is compared to a genomic reference, A-to-G changes can also be a result of single nucleotide polymorphisms (SNPs) present in an individual genome. To limit these errors, methods that involve chemical modification of inosine have also been developed [45, 46]. These approaches have identified a number of ADAR substrates both in *C. elegans* and human brain and were the first to demonstrate that ADAR editing occurs in noncoding regions [47–49]. Comparative genomics was also successful in identifying new ADAR substrates in *Drosophila* and humans [50]. However, the most potent method to identify editing sites has been the use of next generation sequencing.

2.1 Transcriptome-Wide Identification of RNA Editing Sites

When bioinformatics approaches were first utilized to screen human expressed sequence tags (ESTs) for editing events, the number of predicted ADAR substrates skyrocketed from a few dozen to over 1500 mRNAs with over 15,000 editing sites [51–54]. These studies also revealed that the vast majority of A-to-I editing occurs in non-coding regions of mRNAs, such as introns and untranslated regions (UTRs). While these studies suggested widespread effects of ADARs on transcriptome complexity, they all predicted A-to-I editing sites from mismatches between cDNA databases and a reference genome, which increases the possibility that unannotated SNPs are misidentified as editing sites. In fact, after the emergence of next

generation sequencing technology, an elegant study aimed at validating over 36,000 of these bioinformatically predicted editing sites, could only observe a few hundred A-to-I changes present in amplified cDNA and not genomic DNA [55]. This suggests that predicting A-to-I editing sites exclusively from ESTs is not an accurate method of identifying novel ADAR substrates.

More recently, major advances in identifying ADAR substrates have come from the development of stringent bioinformatics pipelines that examine transcriptome-wide RNA-seq data for A-to-I editing events. As A-to-I editing events appear as A-to-G mismatches when comparing mRNA sequencing data to a reference genome (Fig. 8.2), software that allows sequence mismatches in the alignment of the RNA-Seq reads to the genomic DNA is required to successfully map edited reads to the genome. However, the allowance of sequence mismatches also increases the possibility that reads will be misaligned to incorrect genomic loci resulting in the identification of false positive editing sites. In fact, one of the first attempts at identifying editing events in RNA-seq data reported not only A-to-I editing events in the human transcriptome, but also every possible type of nucleotide difference between DNA and RNA [56]. As mechanisms for many of the reported bases changes are not known to exist, these findings were controversial [57]. In addition, other transcriptome-wide studies have suggested that many of these RNA-DNA differences could be attributed to incorrect mapping [58],emphasizing the importance of careful bioinformatics approaches to accurately identify A-to-I editing sites.

To improve the efficacy of bioinformatic pipelines in identifying A-to-I editing sites, a few innovative approaches have been devised [59]. One approach utilizes multiple different mapping tools to stringently map reads containing multiple mismatches, or potential editing events, to one unique location in the genome [34]. A second approach employs a series of filters after mapping, such as requiring a certain minimum number of reads in support of each editing site [60, 61]. A third approach collapses all A and G bases to one base for read alignment then utilizes filters to remove common sources of false positives such as misalignments near exon-exon junctions [35, 62]. A fourth approach has been to sequence RNA from ADAR null organisms and then subtract all the single nucleotide variants present in these transcripts before identification of RNA editing sites [35, 62, 63]. While this approach is limited to model organisms where ADARs are not essential, such as *C. elegans* and *D. melanogaster*, analysis of human cell lines depleted of ADARs has also been used to validate bioinformatically predicted A-to-I editing sites [34]. Together these approaches have expanded the predicted number of editing sites from 10 to 20,000 sites in the first analyses to over one million in recent analyses [60, 61, 64].

Despite the large number of A-to-I editing sites identified in the human transcriptome, the editing sites predicted from separate studies do not significantly overlap. One main reason that editing sites may go undetected is that transcriptome-wide RNA sequencing studies often have low read coverage in dsRNA regions. This low read coverage is primarily a result of inefficient amplification of heavily structured RNA by global reverse transcription techniques. One recent technical advance aimed at comprehensive detection of A-to-I editing sites has been to perform ultra-deep

next generation sequencing (>10,000 read coverage of a given adenosine) of a smaller number of target mRNAs [10]. As the likelihood of detecting low level editing events increases with the depth of read coverage, these approaches are expected to be the most comprehensive method to identifying editing sites for a given gene. By utilizing this technique for Alu repeats in the human transcriptome, every adenosine within these dsRNA forming sequences was reported to be edited, which has raised the potential number of A-to-I events to over 100 million sites in human mRNAs [10]. As many of these sites exhibit editing levels less than 1 %, the biological impact and function of these editing sites is unclear. However, with increased accuracy in detecting editing events, this approach may prove very useful in identifying changes in editing levels between different tissues or healthy and disease states.

3 Regulation of ADAR Editing Activity

Next-generation sequencing has been integral in exploring alterations of A-to-I editing levels during development, in diseases such as brain cancer and in psychiatric disorders [27, 31, 65]. Interestingly, the editing levels do not directly correlate with expression levels of the editing enzymes, underscoring the need to understand how editing is regulated *in vivo*. From the A-to-I regulators that have been identified to date, two general themes for regulating RNA editing levels have emerged: regulating ADAR accessibility to mRNAs and directly altering ADAR function. This chapter will focus on the RNA-binding proteins (RBPs) and other cellular processes that regulate ADAR activity through these general mechanisms.

As ADARs primarily localize to the nucleus and edit pre-mRNAs, alterations in the subcellular localization of ADARs has the potential to alter editing levels in the cell [9]. In humans, ADAR1 and ADAR2 primarily localize to the nucleus, but constantly shuttle in and out of the nucleolus [66, 67]. When shuttling of human ADAR2 to the nucleolus is abrogated, editing levels increase [67]. Therefore, nucleolar sequestration of ADARs is thought to serve as a means to regulate RNA editing. In addition to nucleolar shuttling, human ADAR1 also shuttles from the nucleus to cytoplasm [68]. When ADAR1 is bound to dsRNA, the third dsRBD of ADAR1 adopts a confirmation that promotes shuttling of ADAR1 into the cytoplasm and when dsRNA is absent, ADAR1 is imported into the nucleus [69]. This regulated cycling of ADAR1 is thought to be important for ADAR1 to carry edited RNAs out of the nucleus and prevent re-entry, and perhaps further deamination, of edited RNAs. Although the effect of ADAR1 shuttling on editing efficiency has not been examined, it is presumed that mutations in the third dsRBD of ADAR1, which prevent nuclear re-entry, would lead to decreased editing levels.

Similar to dsRNA-binding for ADAR1, cellular factors from a number of organisms are known to be important for promoting nuclear localization of ADARs and thus increasing editing efficiency [70, 71]. In *C. elegans*, the A-to-I editing enzyme, ADR-2 is mislocalized to the cytoplasm in an *adbp-1* mutant worm line [71].

Although the exact function of ADBP-1 is not known, ADBP-1 localizes exclusively to the nucleus and directly interacts with ADR-2. In addition to mislocalization of ADR-2, worms lacking *adbp-1* have no detectable editing in endogenous mRNAs, suggesting that majority of editing events occur in the nucleus. These data support the hypothesis that ADBP-1 acts as a regulator of RNA editing levels by directing nuclear localization of ADR-2.

While ADBP-1 appears to be a worm specific protein, other proteins, such as Pin1, have been shown to promote nuclear localization of human ADAR2. Pin1 is peptidyl-prolyl *cis*/*trans* isomerase that changes the conformation of prolines in phosphorylated proteins resulting in altered protein function [72]. When Pin1 is knocked down, human ADAR2 is mislocalized to the cytoplasm and editing levels at ADAR2 editing sites decrease [70]. In addition, cells lacking Pin1 have decreased ADAR2 protein stability. The decreased protein stability occurs due to an interaction of ADAR2 with WWP2 in the cytoplasm [70]. WWP2 is an ubiquitin ligase that ubiquitinates ADAR2 leading to its degradation in the cytoplasm. This suggests that Pin1 acts as a global enhancer of ADAR2 editing activity by promoting proper cellular localization, while WWP2 acts as an inhibitor of editing by promoting degradation of human ADAR2. The mechanism for how Pin1 promotes ADAR2 nuclear localization, the kinase that phosphorylates ADAR2 to allow interaction with Pin1, and if this mechanism extends to other ADAR family members are all unknown but promising areas of further research aimed at understanding global regulation of ADAR activity.

Another protein that functions in the ubiquitin degradation pathway, DSS1, was identified in a high throughput screen for regulators of ADAR editing activity [73]. DSS1 acts as an enhancer of ADAR2 editing activity, but the mechanism by which it promotes editing is unknown. DSS1 has been shown to enhance the degradation of ubiquitinated proteins by proteasomes [74, 75]. In addition, DSS1 has been found to localize to the nucleus and enhance protein stability by forming complexes with the DNA repair enzyme BRCA2 and the mRNA exporter TREX-2 [76–78]. These findings have linked DSS1 to a variety of essential cellular processes making functional characterization of DSS1 difficult. DSS1 overexpression increased editing of two mRNAs that contain editing sites within their coding regions, while DSS1 knockdown decreased editing [73]. The C-terminal portion of DSS1 was required for promoting ADAR2 editing. Although this same region is important for the interaction of DSS1 with BRCA2 and TREX-2, a direct interaction between DSS1 and ADAR2 was not detected [76, 78]. An interesting possibility is that DSS1 stabilizes some unidentified ADAR2 enhancer. A likely enhancer candidate is the RNA-binding proteins, hnRNP A2 and hnRNP B1. Both of these proteins physically interact with DSS1 and are known to enhance RNA editing [73]. Future studies aimed at determining the DSS1 interaction partners, including the hnRNPs, are critical next steps to elucidate the regulatory function of DSS1.

In addition to phosphorylation and ubiquitination, A-to-I editing levels can also be affected by sumoylation of ADARs. Human ADAR1 contains a motif for human SUMO-1, which has been shown to sumoylate ADAR1 both *in vitro* and *in vivo*

[79]. Sumoylation of ADAR1 results in decreased editing of dsRNA *in vitro* and a transgenic reporter *in vivo*. However, how sumoylation of ADAR1 alters editing activity and whether endogenous mRNA editing levels are affected is currently unknown. Sumoylation typically disrupts protein-protein interactions, along with changing protein confirmation [80], but whether sumoylated ADAR1 suffers a similar fate remains to be determined.

While many of the identified ADAR regulators directly affect the ADAR proteins, the editing activity of ADAR1 can also be regulated at the level of transcription. In mammals, interferon induces the use of an alternate promoter for the transcription of an ADAR1 isoform, termed p150, which contains Z-DNA binding domains and is largely cytoplasmic [81]. The ADAR1 p150 isoform is thought to have a role in antiviral defense. However, interferon has been linked to changes in editing levels in the human serotonin receptor 5-HT_{2C}R and DNA repair enzyme NEIL1; suggesting that ADAR1 p150 is also capable of regulating editing levels in non-viral editing targets [82, 83]. In contrast to these studies, a recent report on the impact on ADAR1 p150 expression on A-to-I editing levels in mouse brain suggests that induction of this isoform does not significantly affect editing levels of a few known editing sites [84]. However, as neither of the ADAR1 p150 studies examined transcriptome-wide effects on editing, these discrepancies could be influenced by an unknown tissue or species-specific factor. As isoform utilization affects the editing efficiency of other ADAR family members, it is likely that this extends to mammalian ADAR1; however, further study is required to determine the exact roles of the distinct isoforms of ADAR1 and whether their expression correlates to global changes in A-to-I editing.

4 RNA-binding Proteins that Regulate RNA Editing

While many factors directly regulate ADAR editing activity, another major avenue of editing regulation stems from the interaction of RNA-binding proteins (RBPs) with ADAR substrates. From the time that they emerge from RNA polymerase II until they are degraded, mRNAs are coated with RBPs [85]. These RBPs regulate splicing, localization, translation, and stability of the mRNA with the overall outcome determined by the presence of specific binding sites or competition for shared binding locations [86]. In order to edit mRNAs, ADARs must compete with other RBPs for access to dsRNA targets, thus, certain RBPs act as inhibitors of RNA editing. On the other hand, at least one RNA-binding protein, *C. elegans* ADR-1 has been shown to bind to dsRNA and enhance RNA editing [63]. In addition, as RNA editing and splicing both occur co-transcriptionally, RBPs involved in splicing can influence RNA editing [87]. This section will focus on the RBPs that regulate editing in the context of splicing as well as highlight a handful of RBPs that regulate RNA editing through shared targeting of ADAR substrates (Table 8.1).

Table 8.1 RNA-binding proteins (RBPs) that influence A-to-I editing efficiency

RBP regulator	Cellular function	Effect on RNA editing	Organism(s)	Reference
ADARs	A-to-I editing	Repress	Mammals and *Drosophila*	[88, 89]
ADR-1	RNA editing regulator?	Enhance and repress	*C. elegans*	[63]
B52/sRp55	Splicing factor	Repress	*Drosophila*	[90]
DDX15	RNA helicase	Repress	*C. elegans* and humans	[91]
Dicer	miRNA/siRNA processor	Repress	Humans	[92]
FMR-1	Regulation of protein expression	Enhance and repress	*Drosophila*	[93]
hnRNP A2/B1	mRNA transport and processing	Repress	Humans	[73]
Maleless	RNA helicase	Altered	*Drosophila*	[94]
RNA Helicase A	RNA helicase	Repress	Humans	[87]
RPS-14	Ribosomal protein s14	Repress	Humans	[91]
SFRS9	Splicing factor	Repress	Humans	[91]
Staufen	mRNA localization and transport	Unknown	*C. elegans* and humans	[95–97]

RBPs that regulate RNA editing (*column* 1) have a variety of cellular functions (*column* 2), but many are involved in splicing or RNA processing. The expression and/or activity of the RBPs can either enhance or repress editing levels (*column* 3). The organisms in which these regulators have been shown to regulate editing are also listed (*column* 4)

4.1 RNA-binding Proteins that Alter Editing and Splicing

Editing events in coding regions of mRNAs typically involve a dsRNA structure formed between the exon where editing occurs and complementary sequences present in the downstream intron [15]. As the intron required to form the dsRNA would be removed by splicing, splicing can inhibit editing. In addition, secondary structures present in mRNA, such as those targeted by ADARs, influence splicing efficiency [98]. Therefore, these two processes greatly influence each other. This is most evident in the fact that when the length of the dsRNA around the exonic editing site is increased, editing levels increase at the expense of decreased splicing efficiency [99]. Furthermore, when examined *in vitro*, ADARs inhibit proper splicing and, conversely, splicing inhibits ADAR editing activity suggesting that faster splicing would result in less editing [87]. To facilitate editing prior to splicing *in vivo*, ADARs, like other proteins involved in co-transcriptional events, interact with the C-terminal domain of RNA polymerase II (RNAP II) [100, 101]. Removing the C-terminal domain of RNAP II inhibits editing, but promotes splicing of a reporter expressing the exon 11 editing site of the mammalian glutamate receptor mRNA (commonly referred to as the GluR-B Q/R site) [101]. In addition, editing of

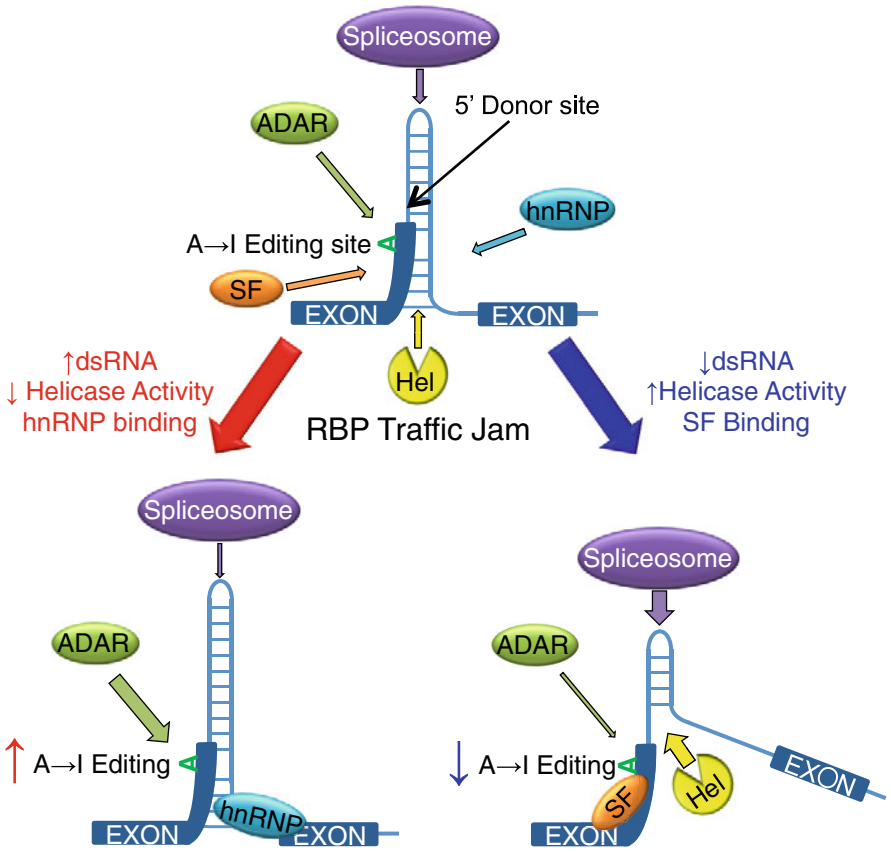


Fig. 8.3 A traffic jam of RBPs occurs at editing sites that are located near splicing signals. Splicing and A-to-I editing are coordinated events, but can interfere with one another due to competition for shared binding sites around the splice junction (*blue hairpin*). (*Lower left*) Editing increases at specific sites when the length of the dsRNA duplex is increased. Binding by hnRNPs (*blue*) and decreased RNA helicase activity (Hel, *yellow*) also result in increased editing levels at the expense of splicing efficiency. (*Lower right*) Mechanisms identified to increase splicing efficiency at the expense of decreased editing levels include competition of splicing factors (*orange*) and ADARs for shared binding sites and increased RNA helicase activity, which unwinds the dsRNA structure around the splice junction

this same reporter promotes splicing, further demonstrating that transcription, editing and splicing are linked events. Because of the complex coordination of these cellular processes, RBPs involved in splicing are promising candidates for regulating A-to-I editing (Fig. 8.3, Table 8.1).

One of the first RBPs found to influence A-to-I editing was identified in a study focused on understanding the complex splicing patterns of the *Drosophila* Na^+ channel gene *para*. Similar to other genes with editing events in codons, *para* contains one exon that base-pairs with complementary sequences located in a downstream

intron. As this double-stranded structure can inhibit splicing, a DEAH box RNA helicase, *maleless*, functions to unwind the dsRNA and promote splicing of *para* transcripts. In fact, loss of *maleless* results in a "splicing catastrophe", with over 80 % of *para* transcripts aberrantly spliced [94]. Interestingly, these splicing defects coincide with altered A-to-I editing levels in *para* suggesting that *maleless* also regulates RNA editing. Consistent with these findings, the mammalian homolog of *maleless*, RNA helicase A (RHA), aids in the proper splicing of the glutamate receptor (GluR-B) pre-mRNA and interferes with the editing activity of mammalian ADAR2, presumably by disrupting the dsRNA structure [87].

Another RNA helicase, mammalian DDX15, was recently identified in a screen for repressors of A-to-I editing [91]. DDX15 repressed editing of substrates that it could directly bind *in vivo*, which may be a general requirement for all helicases that regulate editing. The *C. elegans* DDX15 homolog also repressed A-to-I editing, suggesting that some RBPs regulators of editing may have conserved functions across metazoa. However, as loss of *glh-2*, another *C. elegans* RNA helicase, did not alter editing levels, the ability to regulate editing may not extend to all RNA helicases [91]. Alternatively, if the helicases each have very specific substrates and GLH-2 does not bind the substrate tested, the effects of *glh-2* on editing may not have been detected. In the future, it will be important to conduct transcriptome-wide analyses to determine if RNA helicases act as substrate specific regulators or are capable of regulating A-to-I editing on a global scale.

Although helicases regulate splicing by recognizing and resolving dsRNA structures, most RBPs regulate splicing through sequence specific recognition of mRNA. As A-to-I editing events alter the sequence of the RNA transcript, editing events can affect recognition by RBPs. Consistent with this, a recent transcriptome-wide study determined that A-to-I editing events can alter splicing by creating and destroying splice regulatory elements [102]. Splice regulatory elements are bound by RBPs, such as hnRNPs and SR proteins, that regulate the amount of intron inclusion and exclusion [1]. Specifically, ADAR editing events can create binding sites for the splicing factors SFRS9 and FOX1, while disrupting the binding sites for SFRS5 and hnRNP H1/2 [102]. Interestingly, SFRS9 is also known to bind to ADAR target mRNAs and repress RNA editing [91]. SFRS9 also forms a complex with ADAR2, independent of RNA, suggesting that a direct protein-protein interaction is required for SFRS9 editing regulation. In sum, editing may inhibit or promote recognition by splicing regulatory RBPs; however, these same RBPs can directly regulate editing efficiency, further supporting the idea that that editing and splicing are complexly linked processes.

In addition to editing affecting binding of splicing regulatory factors to pre-mRNAs, RBPs that bind to splicing regulatory elements can alter ADAR accessibility to target mRNAs and thus affect RNA editing levels *in vivo*. One such class of editing regulators is the heterogeneous nuclear ribonucleoproteins (hnRNP) family. This large RBP family contains proteins that regulate intron and exon inclusion and have additional functions in mRNA transport and processing. The wide range of functions for hnRNPs is attributed to the fact that some family members harbor RNA recognition motifs (RRMs) which appear to have both non-specific and

specific preferences for binding mRNA [103]. However, hnRNP family members can also contain KH and Glycine-rich RNA-binding domains to diversify their mRNA-binding targets. Unlike the aforementioned splicing RBPs which all repressed editing levels, knockdown of hnRNP A2/B1 decreases A-to-I editing levels of one specific mRNA, cyFIP2, suggesting that it acts as an enhancer of ADAR editing activity [73]. As hnRNP A2/B1 is known to repress splicing of other mRNAs, the RBP may indirectly increase editing levels by inhibiting the splicing of cyFIP2 and thereby increasing the availability of the dsRNA structure for ADAR2. However, the effects of hnRNP A2/B1 on the splicing efficiency of cyFIP2 are unknown. Furthermore as hnRNP A2/B1 is also known to enhance splicing [104], hnRNP A2/B1 may also indirectly repress A-to-I editing levels in other substrates. Due to the diversity of the hnRNP family and their functions in alternative splicing and other mRNA processing events, it is probable that other unidentified hnRNPs also act as regulators of RNA editing events. Many hnRNP members have known binding sites in RNA, some of which occur within the UTRs of transcripts, suggesting that hnRNPs may also be capable of regulating ADAR editing activity outside of coding regions [103]. Furthermore, using the known hnRNP binding sites, future bioinformatic analyses of ADAR targets for these sequences could result not only in the identification of additional ADAR regulators, but also the specific transcripts they affect.

While RBPs constitute the majority A-to-I editing regulators involved in splicing, small nucleolar RNAs (snoRNAs) can also bind to mRNA and have been implicated in regulating both editing and splicing events. The snoRNA HBII-52 (also called *SNORD115*), which is absent in individuals with Prader-Willi syndrome, promotes the proper splicing of the human serotonin receptor 5-HT_{2C}R, a well-known ADAR target [105–107]. The splice isoform of 5-HT_{2C}R induced by HBII-52 does not contain any detectable editing events suggesting that binding of the snoRNA to the pre-mRNA prevents ADAR access. Consistent with this, expression of the mouse homolog of HBII-52 inhibits A-to-I editing of 5-HT_{2C}R but only in the cellular context of the nucleolus, where ADARs are known to localize [66, 108]. These studies only examined a single gene and snoRNA; therefore, the possibility exists that additional snoRNAs with complementarity to other ADAR targets can also alter editing. In fact, proof of principle studies with synthetic sno-RNAs engineered to bind the mammalian GluR-B transcript inhibited A-to-I editing [108]. Bioinformatic studies examining sequence complementarity between sno-RNAs and ADAR target mRNAs would be useful to understand whether this is a widespread mechanism of regulating A-to-I editing levels.

In addition to specific splicing components regulating RNA editing, a recent study indicates that Period, a protein involved in controlling circadian rhythm, globally affects both alternative splicing and RNA editing [109]. The transcriptional profile of *period* mutant flies express novel splice isoforms of many genes, some of which arise from alternative splicing and many that arise from altered transcriptional start sites. In addition, the *period* mutant flies exhibit altered RNA editing levels, with some sites having increased editing in the absence of *period* and others having decreased editing compared to wild-type. Interestingly, expression of many

of the isoforms and levels of editing were not affected by the time of day, suggesting that Period directly regulates these processes. However, in addition to mRNA transcripts, a number of snoRNAs also had altered expression in the *period* null flies, raising the possibility that *period* affects expression of snoRNAs that directly regulate editing. Interestingly, a long noncoding RNA produced from the same primary transcript as the serotonin editing-regulator, *SNORD115*, was recently shown to regulate expression of circadian clock genes [110]. As this genomic region is silenced in human Prader-Willi syndrome patients, this suggests that, in addition to the snoRNAs from this locus directly regulating editing levels, long non-coding RNAs produced from this locus may regulate circadian rhythm genes and indirectly alter alternative splicing and A-to-I editing.

Last, the most direct role for splicing RBPs in affecting editing levels is the regulated alternative splicing of ADAR transcripts that are known to alter editing activity. In *Drosophila*, the *dADAR* transcript can be alternatively spliced to generate an isoform with additional amino acids between its dsRBDs [88]. This *dADAR* isoform exhibits less editing activity *in vitro* and flies expressing only this isoform have lower levels of RNA editing [111]. The inclusion of this alternate exon is controlled by the splicing factor B52/SRp55 demonstrating that splicing factors can also globally regulate A-to-I editing [90].

In all, these studies demonstrate that ADARs and the RBPs involved in regulating splicing, such as hnRNPs, splicing factors and RNA helicases, are functionally intertwined to carefully regulate both A-to-I editing levels and splicing efficiency of many cellular transcripts.

4.2 Double-Stranded RNA-binding Proteins that Influence RNA Editing

As most dsRNA-binding proteins (dsRBPs), including ADARs, utilize the dsRNA-binding domain (dsRBD) to recognize dsRNA in a structure specific, sequence-independent manner and this domain is second-most abundant RNA recognition motif, dsRBPs make up a large number of editing regulators. The first and foremost dsRBP regulators of editing are members of the ADAR family itself. ADARs affect global A-to-I editing levels through self-editing of ADAR transcripts that give rise to ADAR isoforms with different editing activities. In both mammalian ADAR2 and *Drosophila* ADAR, the pre-mRNA transcripts contain complementary elements that form dsRNA and can be edited to give rise to an ADAR protein that is less active at editing [88, 89]. However, the editing event in mammalian ADAR2 causes an alternative splicing event resulting in a truncated protein with no editing function, while the editing event in *Drosophila* ADAR mutates an essential amino acid in the catalytic domain to reduce editing activity. As mice unable to self-edit have increased ADAR2 protein levels, which can lead to obesity, and flies lacking ability to self-editing are not viable, these

self-editing mechanisms provide critical negative feedback to control ADAR editing levels [111–113].

In addition to ADAR self-editing activity regulating A-to-I editing levels, the dsRBDs of ADARs can also influence editing activity. When examining ADAR2 editing activity *in vitro*, the full length protein was found to be incapable of editing a small 15 bp dsRNA substrate [114]. However, an ADAR2 construct lacking the N-terminal dsRBD could efficiently edit this small substrate, suggesting that the N-terminal dsRBD of human ADAR2 acts to inhibit editing of specific substrates. Furthermore, as this effect is only detected with small substrates that cannot accommodate binding sites for both dsRBDs of human ADAR2, it is proposed that protein-protein interactions between the two dsRBDs leads to inhibition of editing activity. Whether this auto-inhibition activity extends to other ADARs with multiple dsRBDs is unknown. However, in other dsRBPs that contain multiple dsRBDs, such as Staufen, certain dsRBDs have evolved specialized functions beyond binding dsRNA [115].

ADAR family members are designated by the homology shared in their conserved catalytic domain and the presence of N-terminal dsRBDs; however, not all ADAR family members can edit dsRNA. For example, in *C. elegans*, two ADAR family members, ADR-1 and ADR-2, are encoded in the genome [116]. However, worms lacking *adr-2*, but expressing *adr-1*, lack A-to-I editing of endogenous mRNAs, suggesting that ADR-1 is an inactive deaminase [23]. Interestingly, worms lacking *adr-1* exhibit altered editing levels [23]. A recent study demonstrated that ADR-1 interacts with the editing targets of ADR-2, but does not physically interact with ADR-2 in the absence of dsRNA [63]. Furthermore, mutations in the dsRBDs of ADR-1 that disrupt binding to mRNA *in vivo*, abrogated the ability of ADR-1 to regulate editing levels. Interestingly, ADR-1 primarily promotes editing and appears to regulate A-to-I editing events across the transcriptome, which has not been reported for any other known A-to-I editing regulators. Although the exact mechanism for how ADR-1 regulates editing is unknown, the binding of the dsRBD of another dsRBP, yeast Rnt1p, to dsRNA has recently been shown to alter the structure of dsRNA [117]; therefore, an interesting possibility is that ADR-1 binding to target mRNA alters the dsRNA structure surrounding editing sites thereby increasing the catalytic activity of ADR-2 for the target adenosine or increasing the affinity of ADR-2 for the target dsRNA.

Deaminase-deficient ADAR family members are not limited to nematodes. In mammals, two ADAR family members, TENR and ADAR3, are also thought to lack deaminase activity. TENR, which is a testis specific ADAR family member, is most homologous to *C. elegans* ADR-1 [32]. Mice lacking TENR expression have sperm defects and are sterile, suggesting TENR plays a critical role in spermatogenesis [118]. An intriguing possibility is that TENR regulates RNA editing of mRNAs required for spermatogenesis.

The other mammalian ADAR predicted to lack RNA editing activity is the brain-specific protein, ADAR3 [119]. Recombinant, purified ADAR3 did not exhibit editing activity *in vitro* [120]. However, as ADAR3 does possess the known amino acids

required for catalysis, it is unclear why it lacks editing activity. Despite an apparent lack of editing activity, ADAR3 inhibits editing by both ADAR1 and ADAR2 *in vitro*, suggesting that it could serve as a regulator of A-to-I editing *in vivo* [120]. However, as ADAR1 and ADAR2 also inhibit each other's editing activity in this *in vitro* assay, the inhibitory activity may simply be a result of competition for substrates that occurs in a test tube, but not in the cellular environment.

Another dsRBP that has recently been shown to regulate A-to-I editing levels is the RNaseIII enzyme, Dicer [92]. Dicer is required for processing dsRNA molecules to give rise to both mature siRNAs and miRNAs [121]. Co-immunoprecipitation experiments revealed that Dicer interacts directly with human ADAR1 [92]. This result was surprising as Dicer is thought to primarily localize to the cytoplasm, whereas ADAR1 edits pre-mRNAs in the nucleus. The heterodimer of Dicer and ADAR1 has an increased rate of miRNA processing compared to Dicer alone [92]. However, when ADAR1 was in complex with Dicer, ADAR1 editing activity was inhibited *in vitro*, suggesting that Dicer acts as an inhibitor of editing. It is presumed that this inhibition arises due to the sequestration of ADAR1 from editing targets. However, the exact mechanism that Dicer utilizes to regulate editing is unknown. Interestingly, ADAR1 can also inhibit the expression of Dicer through regulation of let-7 miRNAs suggesting that Dicer and ADAR1 are involved in a negative feedback loop [122]. Whether increases in Dicer protein expression result in decreases in A-to-I editing levels *in vivo* is unknown, but is an interesting area of further research.

Another dsRBP, Staufen, has recently been shown to associate with ADAR target mRNAs in multiple organisms; however, a role for Staufen in regulating editing is currently unclear. Staufen is involved in many post-transcriptional processes, including regulating mRNA localization, translation and decay [115, 123]. A study of *C. elegans* STAU-1 bound mRNAs identified several genes that are edited by *C. elegans* ADR-2 [63, 97]. In addition, a study of the RNA targets of human Staufen1 detected numerous inverted Alu repeat elements in mRNAs, which are also known to be human ADAR targets [60, 95]. Neither of the Staufen studies directly determined whether Staufen binding to ADAR target mRNAs impacted ADAR accessibility for the dsRNA. However, the overall number of editing events within a reporter harboring inverted Alu repeat elements was examined and reported to be unaffected by alterations in human Staufen1 expression [96]. However, to the reader, these reporters appear to show alterations of editing levels at individual sites when comparing wild-type cells and those with decreased Staufen levels. While statistical differences in editing levels at individual adenosines needs to be examined to confirm this observation, this suggests that Staufen may be capable of regulating editing levels *in vivo*. Furthermore, as the effect of Staufen on editing has only been analyzed for a reporter mRNA, the possibility exists that Staufen1 alters editing of endogenous mRNAs that contain Alu elements. As many of the aforementioned RBPs regulate A-to-I editing levels by binding to the editing targets of ADARs, Staufen appears as a promising candidate to act as a conserved regulator of A-to-I editing.

4.3 Disease-Associated RNA-binding Proteins that Regulate RNA Editing

While splicing components and dsRBPs constitute the majority of RBPs that regulate A-to-I editing levels, two other well-known RBPs, FMR-1 and RPS-14, have recently been shown to alter editing levels [91, 93]. FMR-1 is an RBP containing both KH and RGG box RNA-binding domains [124]. The KH domains of FMR-1 are required for its *in vivo* function as point mutations within these domains disrupts RNA-binding and are present in patients suffering from fragile X syndrome [125–127]. Flies lacking the FMR-1 homolog, dFMR1, exhibit altered RNA editing levels in five of six ADAR targets tested [93]. In addition, flies overexpressing dFMR1 lead to differential effects on editing; however, only two of the editing sites demonstrated a clear bi-directional change in response to reduced or increased dFMR1 expression. dFMR1 binds to the ADAR target mRNAs *in vivo*, and mutations to both KH domains result in altered editing levels in two endogenous ADAR targets when compared to wild-type. These data suggest that FMR-1 regulates RNA editing by acting as a RBP. However, it is unknown whether RNA-binding by the KH domains is required for FMR-1 to regulate editing of other target mRNAs or perhaps RNA-binding by the RGG box domain also plays a role in regulating editing levels. Given these results, it is likely that FMR-1 has an additional cellular role beyond its well-studied function in regulating protein expression [128]. It would be interesting to explore if patients with Fragile X syndrome also exhibit altered A-to-I editing levels similar to the *Drosophila* model.

The 40S ribosomal subunit protein S14 (RPS-14) represses ADAR editing *in vivo* in a substrate dependent manner [91]. RPS-14 is an accessory protein that binds directly to the 18S rRNA and is required for the assembly of the small ribosomal subunit [129]. Similar to SFRS9, mammalian RPS-14 was found to co-immunoprecipitate with rat ADAR2, irrespective of treatment with RNase [91]. Therefore, it is possible that heterodimerization of ADAR2 with RPS14 results in repression of editing. However, as RPS-14 only repressed editing of the cyFIP2 mRNA, which it also directly bound, RNA-binding by RPS-14 is likely a critical determining factor for its regulatory ability. As RPS-14 localizes solely to the nucleolus, it is possible that RPS-14 regulates ADAR2 by sequestering it in the nucleolus or that it can only regulate editing levels in transcripts that localize to the nucleolus. In addition to ensuring proper ribosome maturation, RPS14 is required for proper production of differentiated erythroid cells and loss of RPS14 occurs in a human myelodysplastic syndrome, known as 5q- syndrome [130]. Similarly, ADAR1 is reported to be essential for hematopoiesis, with mice lacking ADAR1 showing severe defects in proliferation and/or differentiation of erythrocytes [21, 131]. While no interaction between RPS-14 and ADAR1 has been reported, the roles of both proteins in promoting efficient erythropoiesis suggest that RPS-14 may also regulate similar targets as ADAR1. As other editing regulators localize exclusively to the nucleolus, which is also a site of ADAR localization, it is very likely RPS-14 represents the first of many nucleolar RBPs that serve as ADAR editing regulators.

Furthermore, as loss of both ADARs and several ribosomal RNA-binding proteins are implicated in hematopoietic disorders, transcriptome-wide analysis of A-to-I editing levels in patients with 5q- syndrome or other ribosomopathies is an important first step in understanding the interplay amongst these RBP pathways.

5 Conclusions

With over 60 % of human transcripts predicted to contain adenosine to inosine modifications, RNA editing plays a major role in generating molecular complexity. However, because of the potential of A-to-I editing to alter the flow of genetic information, ADAR editing activity must also be highly regulated. Genetic and biochemical approaches have identified RNA-binding proteins (RBPs) that regulate editing. The majority of these RBPs were discovered due to their ability to regulate editing of a specific adenosine within a target mRNA. However, as most of the regulators identified across all organisms fall into three main classes of RBPs: splicing factors, RNA helicases and dsRNA-binding proteins, this suggests that there are conserved mechanisms that function to regulate editing *in vivo*. Future work aimed at dissecting these mechanisms not only in normal cells, but also in the multiple neurological and cancerous tissues that exhibit altered editing is critical. In addition, the use of stringent bioinformatics and next-generation sequencing technology is essential for determining the global impact of many of the ADAR editing regulators on both RNA editing of transcripts and gene expression. Finally, as many of the RBPs that regulate ADAR editing are part of larger functional classes, it is likely that additional RBPs will be identified as important cellular regulators of A-to-I editing.

Acknowledgement This work was supported by an NIH predoctoral training grant to M.C.W. (T32 GM007757) and start-up funds from the Indiana University School of Medicine.

References

1. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10(11):741–754. doi:10.1038/nrm2777
2. Farajollahi S, Maas S (2010) Molecular diversity through RNA editing: a balancing act. Trends Genet 26(5):221–230. doi:10.1016/j.tig.2010.02.001
3. Tariq A, Jantsch MF (2012) Transcript diversification in the nervous system: a to I RNA editing in CNS function and disease development. Front Neurosci 6:99. doi:10.3389/fnins.2012.00099
4. Bass BL, Weintraub H (1988) An unwinding activity that covalently modifies its double-stranded-RNA substrate. Cell 55(6):1089–1098. doi:10.1016/0092-8674(88)90253-x
5. Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC (1986) Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46(6):819–826. doi:10.1016/0092-8674(86)90063-2

6. Gott JM, Emeson RB (2000) Functions and mechanisms of RNA editing. Annu Rev Genet 34(1):499–531. doi:10.1146/annurev.genet.34.1.499
7. Keegan LP, Gallo A, O'Connell MA (2001) The many roles of an RNA editor. Nat Rev Genet 2(11):869–878
8. Blanc V, Davidson NO (2010) APOBEC-1-mediated RNA editing. Wiley Interdiscip Rev Syst Biol Med 2(5):594–602. doi:10.1002/wsbm.82
9. Nishikura K (2010) Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 79:321–349. doi:10.1146/annurev-biochem-060208-105251
10. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P, Zhang R, Isaacs FJ, Rechavi G, Li JB, Eisenberg E, Levanon EY (2014) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res 24(3):365–376. doi:10.1101/gr.164749.113
11. Blanc V, Park E, Schaefer S, Miller M, Lin Y, Kennedy S, Billing A, Hamidane H, Graumann J, Mortazavi A, Nadeau J, Davidson N (2014) Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. Genome Biol 15(6):R79
12. Hamilton CE, Papavasiliou FN, Rosenberg BR (2010) Diverse functions for DNA and RNA editing in the immune system. RNA Biol 7(2):220–228
13. Savva YA, Rieder LE, Reenan RA (2012) The ADAR protein family. Genome Biol 13(12):252. doi:10.1186/gb-2012-13-12-252
14. Rosenthal JJ, Seeburg PH (2012) A-to-I RNA editing: effects on proteins key to neural excitability. Neuron 74(3):432–439. doi:10.1016/j.neuron.2012.04.010
15. Rieder LE, Reenan RA (2012) The intricate relationship between RNA structure, editing, and splicing. Semin Cell Dev Biol 23(3):281–288. doi:10.1016/j.semcdb.2011.11.004
16. Tomaselli S, Bonamassa B, Alisi A, Nobili V, Locatelli F, Gallo A (2013) ADAR enzyme and miRNA story: a nucleotide that can make the difference. Int J Mol Sci 14(11):22796–22816. doi:10.3390/ijms141122796
17. Hundley HA, Bass BL (2010) ADAR editing in double-stranded UTRs and other noncoding RNA sequences. Trends Biochem Sci 35(7):377–383. doi:10.1016/j.tibs.2010.02.008
18. Wahlstedt H, Ohman M (2011) Site-selective versus promiscuous A-to-I editing. Wiley Interdiscip Rev RNA 2(6):761–771. doi:10.1002/wrna.89
19. Wulff BE, Nishikura K (2012) Modulation of microRNA expression and function by ADARs. Curr Top Microbiol Immunol 353:91–109. doi:10.1007/82_2011_151
20. Higuchi M, Single FN, Kohler M, Sommer B, Sprengel R, Seeburg PH (1993) RNA editing of AMPA receptor subunit Glur-B—A base-paired intron-exon structure determines position and efficiency. Cell 75(7):1361–1370. doi:10.1016/0092-8674(93)90622-w
21. Wang Q, Khillan J, Gadue P, Nishikura K (2000) Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. Science 290(5497):1765–1768. doi:10.1126/science.290.5497.1765
22. 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(4):437–449. doi:10.1016/S0092-8674(00)00049-0
23. 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(22):6025–6035
24. Gaisler-Salomon I, Kravitz E, Feiler Y, Safran M, Biegon A, Amariglio N, Rechavi G (2014) Hippocampus-specific deficiency in RNA editing of GluA2 in Alzheimer's disease. Neurobiol Aging. doi:10.1016/j.neurobiolaging.2014.02.018
25. Galeano F, Rossetti C, Tomaselli S, Cifaldi L, Lezzerini M, Pezzullo M, Boldrini R, Massimi L, Di Rocco CM, Locatelli F, Gallo A (2013) ADAR2-editing activity inhibits glioblastoma growth through the modulation of the CDC14B/Skp2/p21/p27 axis. Oncogene 32(8):998–1009. doi:10.1038/onc.2012
26. Krestel H, Raffel S, von Lehe M, Jagella C, Moskau-Hartmann S, Becker A, Elger CE, Seeburg PH, Nirkko A (2013) Differences between RNA and DNA due to RNA editing in temporal lobe epilepsy. Neurobiol Dis 56:66–73. doi:10.1016/j.nbd.2013.04.006

27. Silberberg G, Lundin D, Navon R, Öhman M (2011) Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. Hum Mol Genet 21(2):311–321. doi:10.1093/hmg/ddr461
28. Yamashita T, Kwak S (2013) The molecular link between inefficient GluA2 Q/R site-RNA editing and TDP-43 pathology in motor neurons of sporadic amyotrophic lateral sclerosis patients. Brain Res 1584:28–38. doi:10.1016/j.brainres.2013.12.011
29. Hideyama T, Yamashita T, Aizawa H, Tsuji S, Kakita A, Takahashi H, Kwak S (2012) Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. Neurobiol Dis 45(3):1121–1128. doi:10.1016/j.nbd.2011.12.033
30. Maas S, Patt S, Schrey M, Rich A (2001) Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proc Natl Acad Sci U S A 98(25):14687–14692. doi:10.1073/pnas.251531398
31. Wahlstedt H, Daniel C, Enstërö M, Öhman M (2009) Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res 19(6):978–986. doi:10.1101/gr.089409.108
32. Jin YF, Zhang WJ, Li Q (2009) Origins and evolution of ADAR-mediated RNA editing. IUBMB Life 61(6):572–578. doi:10.1002/iub.207
33. Lehmann KA, Bass BL (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39(42):12875–12884. doi:10.1021/bi001383g
34. Bahn JH, Lee JH, Li G, Greer C, Peng G, Xiao X (2012) Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res 22(1):142–150. doi:10.1101/gr.124107.111
35. St Laurent G, Tackett MR, Nechkin S, Shtokalo D, Antonets D, Savva YA, Maloney R, Kapranov P, Lawrence CE, Reenan RA (2013) Genome-wide analysis of A-to-I RNA editing by single-molecule sequencing in Drosophila. Nat Struct Mol Biol 20(11):1333–1339. doi:10.1038/nsmb.2675
36. Wong SK, Sato S, Lazinski DW (2001) Substrate recognition by ADAR1 and ADAR2. RNA 7(6):846–858. doi:10.1017/s135583820101007x
37. Eggington JM, Greene T, Bass BL (2011) Predicting sites of ADAR editing in double-stranded RNA. Nat Commun 2:319. doi:10.1038/ncomms1324
38. Kuttan A, Bass BL (2012) Mechanistic insights into editing-site specificity of ADARs. Proc Natl Acad Sci U S A 109(48):E3295–E3304. doi:10.1073/pnas.1212548109
39. Goodman RA, Macbeth MR, Beal PA (2012) ADAR proteins: structure and catalytic mechanism. Curr Top Microbiol Immunol 353:1–33. doi:10.1007/82_2011_144
40. Liu Y, Lei M, Samuel CE (2000) Chimeric double-stranded RNA-specific adenosine deaminase ADAR1 proteins reveal functional selectivity of double-stranded RNA-binding domains from ADAR1 and protein kinase PKR. Proc Natl Acad Sci U S A 97(23):12541–12546. doi:10.1073/pnas.97.23.12541
41. Barraud P, Heale BS, O'Connell MA, Allain FH (2012) Solution structure of the N-terminal dsRBD of Drosophila ADAR and interaction studies with RNA. Biochimie 94(7):1499–1509. doi:10.1016/j.biochi.2011.12.017
42. Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB, Allain FH (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell 143(2):225–237. doi:10.1016/j.cell.2010.09.026
43. Eisenberg E (2012) Bioinformatic approaches for identification of A-to-I editing sites. Curr Top Microbiol Immunol 353:145–162. doi:10.1007/82_2011_147
44. Sommer B, Kohler M, Sprengel R, Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell 67(1):11–19. doi:10.1016/0092-8674(91)90568-j
45. Morse DP, Bass BL (1997) Detection of inosine in messenger RNA by inosine-specific cleavage. Biochemistry 36(28):8429–8434. doi:10.1021/bi9709607

46. Sakurai M, Yano T, Kawabata H, Ueda H, Suzuki T (2010) Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. Nat Chem Biol 6(10):733–740. doi:10.1038/nchembio.434
47. Morse DP, Aruscavage PJ, Bass BL (2002) RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. Proc Natl Acad Sci U S A 99(12):7906–7911. doi:10.1073/pnas.112704299
48. Morse DP, Bass BL (1999) Long RNA hairpins that contain inosine are present in Caenorhabditis elegans poly(A)+RNA. Proc Natl Acad Sci 96(11):6048–6053. doi:10.1073/pnas.96.11.6048
49. Sakurai M, Ueda H, Yano T, Okada S, Terajima H, Mitsuyama T, Toyoda A, Fujiyama A, Kawabata H, Suzuki T (2014) A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. Genome Res 24(3):522–534. doi:10.1101/gr.162537.113
50. Hoopengardner B, Bhalla T, Staber C, Reenan R (2003) Nervous system targets of RNA editing identified by comparative genomics. Science 301(5634):832–836. doi:10.1126/science.1086763
51. Athanasiadis A, Rich A, Maas S (2004) Widespread A-to-I RNA editing of alu-containing mRNAs in the human transcriptome. PLoS Biol 2(12):2144–2158. doi:10.1371/journal.pbio.0020391
52. Blow M, Futreal PA, Wooster R, Stratton MR (2004) A survey of RNA editing in human brain. Genome Res 14(12):2379–2387. doi:10.1101/gr.2951204
53. Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A (2004) Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res 14(9):1719–1725. doi:10.1101/gr.2855504
54. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, Olshansky M, Rechavi G, Jantsch MF (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol 22(8):1001–1005. doi:10.1038/nbt996
55. Li JB, Levanon EY, Yoon J-K, Aach J, Xie B, LeProust E, Zhang K, Gao Y, Church GM (2009) Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 324(5931):1210–1213. doi:10.1126/science.1170995
56. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG (2011) Widespread RNA and DNA sequence differences in the human transcriptome. Science 333(6038):53–58. doi:10.1126/science.1207018
57. Schrider DR, Gout J-F, Hahn MW (2011) Very few RNA and DNA sequence differences in the human transcriptome. PLoS One 6(10), e25842. doi:10.1371/journal.pone.0025842
58. Piskol R, Peng Z, Wang J, Li JB (2013) Lack of evidence for existence of noncanonical RNA editing. Nat Biotechnol 31(1):19–20. doi:10.1038/nbt.2472
59. Lee JH, Ang JK, Xiao X (2013) Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. RNA. doi:10.1261/rna.037903.112
60. 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(6):579–581. doi:10.1038/nmeth.1982
61. Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, Guo J, Dong Z, Bao L, Wang J (2012) Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat Biotechnol 30(3):253–260. doi:10.1038/nbt.2122
62. Wu D, Lamm AT, Fire AZ (2011) Competition between ADAR and RNAi pathways for an extensive class of RNA targets. Nat Struct Mol Biol 18(10):1094–1101. doi:10.1038/Nsmb.2129
63. Washburn MC, Kakaradov B, Sundararaman B, Wheeler E, Hoon S, Yeo GW, Hundley HA (2014) The dsRBP and inactive editor ADR-1 utilizes dsRNA-binding to regulate A-to-I RNA editing across the C. elegans transcriptome. Cell Rep 6(4):599–607. doi:10.1016/j.celrep.2014.01.011

64. Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB (2013) Identifying RNA editing sites using RNA sequencing data alone. Nat Methods 10(2):128–132. doi:10.1038/nmeth.2330
65. Paz N, Levanon EY, Amariglio N, Heimberger AB, Ram Z, Constantini S, Barbash ZS, Adamsky K, Safran M, Hirschberg A, Krupsky M, Ben-Dov I, Cazacu S, Mikkelsen T, Brodie C, Eisenberg E, Rechavi G (2007) Altered adenosine-to-inosine RNA editing in human cancer. Genome Res 17(11):1586–1595. doi:10.1101/gr.6493107
66. Desterro JMP, Keegan LP, Lafarga M, Berciano MT, O'Connell M, Carmo-Fonseca M (2003) Dynamic association of RNA-editing enzymes with the nucleolus. J Cell Sci 116(9):1805–1818. doi:10.1242/jcs.00371
67. Sansam CL, Wells KS, Emeson RB (2003) Modulation of RNA editing by functional nucleolar sequestration of ADAR2. Proc Natl Acad Sci 100(24):14018–14023. doi:10.1073/pnas.2336131100
68. Fritz J, Strehblow A, Taschner A, Schopoff S, Pasierbek P, Jantsch MF (2009) RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. Mol Cell Biol 29(6):1487–1497. doi:10.1128/MCB.01519-08
69. Barraud P, Banerjee S, Mohamed WI, Jantsch MF, Allain FH (2014) A bimodular nuclear localization signal assembled via an extended double-stranded RNA-binding domain acts as an RNA-sensing signal for transportin 1. Proc Natl Acad Sci U S A 111(18):E1852–E1861. doi:10.1073/pnas.1323698111
70. Marcucci R, Brindle J, Paro S, Casadio A, Hempel S, Morrice N, Bisso A, Keegan LP, Del Sal G, O'Connell MA (2011) Pin1 and WWP2 regulate GluR2 Q/R site RNA editing by ADAR2 with opposing effects. EMBO J 30(20):4211–4222. doi:10.1038/emboj.2011.303
71. Ohta H, Fujiwara M, Ohshima Y, Ishihara T (2008) ADBP-1 regulates an ADAR RNA-editing enzyme to antagonize RNA-interference-mediated gene silencing in Caenorhabditis elegans. Genetics 180(2):785–796. doi:10.1534/genetics.108.093310
72. Lu KP, Zhou XZ (2007) The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. Nat Rev Mol Cell Biol 8(11):904–916. doi:10.1038/nrm2261
73. Garrncarz W, Tariq A, Handl C, Pusch O, Jantsch MF (2013) A high throughput screen to identify enhancers of ADAR-mediated RNA-editing. RNA Biol 10(2):192–204. doi:10.4161/rna.23208
74. Funakoshi M, Li X, Velichutina I, Hochstrasser M, Kobayashi H (2004) Sem1, the yeast ortholog of a human BRCA2-binding protein, is a component of the proteasome regulatory particle that enhances proteasome stability. J Cell Sci 117(26):6447–6454. doi:10.1242/jcs.01575
75. Wei SJ, Williams JG, Dang H, Darden TA, Betz BL, Humble MM, Chang FM, Trempus CS, Johnson K, Cannon RE, Tennant RW (2008) Identification of a specific motif of the DSS1 protein required for proteasome interaction and p53 protein degradation. J Mol Biol 383(3):693–712. doi:10.1016/j.jmb.2008.08.044
76. Ellisdon AM, Dimitrova L, Hurt E, Stewart M (2012) Structural basis for the assembly and nucleic acid binding of the TREX-2 transcription-export complex. Nat Struct Mol Biol 19(3):328–336. doi:10.1038/nsmb.2235
77. Li J, Zou C, Bai Y, Wazer DE, Band V, Gao Q (2005) DSS1 is required for the stability of BRCA2. Oncogene 25(8):1186–1194
78. Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thomä NH, Zheng N, Chen P-L, Lee W-H, Pavletich NP (2002) BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science 297(5588):1837–1848. doi:10.1126/science.297.5588.1837
79. Desterro JM, Keegan LP, Jaffray E, Hay RT, O'Connell MA, Carmo-Fonseca M (2005) SUMO-1 modification alters ADAR1 editing activity. Mol Biol Cell 16(11):5115–5126. doi:10.1091/mbc.E05-06-0536
80. Wilkinson KA, Henley JM (2010) Mechanisms, regulation and consequences of protein SUMOylation. Biochem J 428(2):133–145. doi:10.1042/bj20100158

81. George CX, Gan Z, Liu Y, Samuel CE (2011) Adenosine deaminases acting on RNA, RNA editing, and interferon action. J Interferon Cytokine Res 31(1):99–117. doi:10.1089/jir.2010.0097
82. Yang W, Wang Q, Kanes SJ, Murray JM, Nishikura K (2004) Altered RNA editing of serotonin 5-HT2C receptor induced by interferon: implications for depression associated with cytokine therapy. Brain Res Mol Brain Res 124(1):70–78. doi:10.1016/j.molbrainres.2004.02.010
83. Yeo J, Goodman RA, Schirle NT, David SS, Beal PA (2010) RNA editing changes the lesion specificity for the DNA repair enzyme NEIL1. Proc Natl Acad Sci 107(48):20715–20719. doi:10.1073/pnas.1009231107
84. Hood JL, Morabito MV, Martinez CR, Gilbert JA, Ferrick EA, Ayers GD, Chappell JD, Dermody TS, Emeson RB (2014) Reovirus-mediated induction of ADAR1 (p150) minimally alters RNA editing patterns in discrete brain regions. Mol Cell Neurosci. doi:10.1016/j.mcn.2014.06.001
85. Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science 309(5740):1514–1518. doi:10.1126/science.1111443
86. Anderson P, Kedersha N (2009) RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat Rev Mol Cell Biol 10(6):430–436. doi:10.1038/nrm2694
87. Bratt E, Ohman M (2003) Coordination of editing and splicing of glutamate receptor pre-mRNA. RNA 9(3):309–318. doi:10.1261/rna.2750803
88. Palladino MJ, Keegan LP, O'Connell MA, Reenan RA (2000) dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. RNA 6(7):1004–1018
89. Rueter SM, Dawson TR, Emeson RB (1999) Regulation of alternative splicing by RNA editing. Nature 399(6731):75–80
90. Marcucci R, Romano M, Feiguin F, O'Connell MA, Baralle FE (2009) Dissecting the splicing mechanism of the Drosophila editing enzyme; dADAR. Nucleic Acids Res 37(5):1663–1671. doi:10.1093/nar/gkn1080
91. Tariq A, Garncarz W, Handl C, Balik A, Pusch O, Jantsch MF (2013) RNA-interacting proteins act as site-specific repressors of ADAR2-mediated RNA editing and fluctuate upon neuronal stimulation. Nucleic Acids Res 41(4):2581–2593. doi:10.1093/nar/gks1353
92. Ota H, Sakurai M, Gupta R, Valente L, Wulff B-E, Ariyoshi K, Iizasa H, Davuluri Ramana V, Nishikura K (2013) ADAR1 forms a complex with dicer to promote MicroRNA processing and RNA-induced gene silencing. Cell 153(3):575–589. doi:10.1016/j.cell.2013.03.024
93. Bhogal B, Jepson JE, Savva YA, Pepper AS, Reenan RA, Jongens TA (2011) Modulation of dADAR-dependent RNA editing by the Drosophila fragile X mental retardation protein. Nat Neurosci 14(12):1517–1524. doi:10.1038/nn.2950
94. Reenan RA, Hanrahan CJ, Ganetzky B (2000) The mlenapts RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+channel transcript in a region of RNA editing. Neuron 25(1):139–149. doi:10.1016/s0896-6273(00)80878-8
95. de Lucas S, Oliveros JC, Chagoyen M, Ortin J (2014) Functional signature for the recognition of specific target mRNAs by human Staufen1 protein. Nucleic Acids Res. doi:10.1093/nar/gku073
96. Elbarbary RA, Li W, Tian B, Maquat LE (2013) STAU1 binding 3′ UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. Genes Dev 27(13):1495–1510. doi:10.1101/gad.220962.113
97. Legendre JB, Campbell ZT, Kroll-Conner P, Anderson P, Kimble J, Wickens M (2013) RNA targets and specificity of Staufen, a double-stranded RNA-binding protein in Caenorhabditis elegans. J Biol Chem 288(4):2532–2545. doi:10.1074/jbc.M112.397349
98. Buratti E, Baralle FE (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. Mol Cell Biol 24(24):10505–10514. doi:10.1128/MCB.24.24.10505-10514.2004
99. Rieder LE, Staber CJ, Hoopengardner B, Reenan RA (2013) Tertiary structural elements determine the extent and specificity of messenger RNA editing. Nat Commun 4:2232. doi:10.1038/ncomms3232

100. Laurencikiene J, Kallman AM, Fong N, Bentley DL, Ohman M (2006) RNA editing and alternative splicing: the importance of co-transcriptional coordination. EMBO Rep 7(3):303–307. doi:10.1038/sj.embor.7400621
101. Ryman K, Fong N, Bratt E, Bentley DL, Ohman M (2007) The C-terminal domain of RNA Pol II helps ensure that editing precedes splicing of the GluR-B transcript. RNA 13(7):1071–1078. doi:10.1261/rna.404407
102. Solomon O, Oren S, Safran M, Deshet-Unger N, Akiva P, Jacob-Hirsch J, Cesarkas K, Kabesa R, Amariglio N, Unger R, Rechavi G, Eyal E (2013) Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). RNA 19(5):591–604. doi:10.1261/rna.038042.112
103. Han SP, Tang YH, Smith R (2010) Functional diversity of the hnRNPs: past, present and perspectives. Biochem J 430(3):379–392. doi:10.1042/BJ20100396
104. Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, Revil T, Chabot B (2007) hnRNP proteins and splicing control. Adv Exp Med Biol 623:123–147
105. Burns CM, Chu H, Rueter SM, Hutchinson LK, Canton H, Sanders-Bush E, Emeson RB (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature 387(6630):303–308. doi:10.1038/387303a0
106. Cavaillé J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie J-P, Brosius J, Hüttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc Natl Acad Sci 97(26):14311–14316. doi:10.1073/pnas.250426397
107. Kishore S, Stamm S (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. Science 311(5758):230–232. doi:10.1126/science.1118265
108. Vitali P, Basyuk E, Le Meur E, Bertrand E, Muscatelli F, Cavaille J, Huttenhofer A (2005) ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. J Cell Biol 169(5):745–753. doi:10.1083/jcb.200411129
109. Hughes ME, Grant GR, Paquin C, Qian J, Nitabach MN (2012) Deep sequencing the circadian and diurnal transcriptome of Drosophila brain. Genome Res 22(7):1266–1281. doi:10.1101/gr.128876.111
110. Powell WT, Coulson RL, Crary FK, Wong SS, Ach RA, Tsang P, Alice Yamada N, Yasui DH, Lasalle JM (2013) A Prader-Willi locus lncRNA cloud modulates diurnal genes and energy expenditure. Hum Mol Genet 22(21):4318–4328. doi:10.1093/hmg/ddt281
111. Keegan LP, Brindle J, Gallo A, Leroy A, Reenan RA, Connell MA (2005) Tuning of RNA editing by ADAR is required in Drosophila. EMBO J 24(12):2183–2193
112. Feng Y, Sansam CL, Singh M, Emeson RB (2006) Altered RNA editing in mice lacking ADAR2 autoregulation. Mol Cell Biol 26(2):480–488. doi:10.1128/MCB.26.2.480-488.2006
113. Singh M, Kesterson RA, Jacobs MM, Joers JM, Gore JC, Emeson RB (2007) Hyperphagia-mediated obesity in transgenic mice misexpressing the RNA-editing enzyme ADAR2. J Biol Chem 282(31):22448–22459. doi:10.1074/jbc.M700265200
114. Macbeth MR, Lingam AT, Bass BL (2004) Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA-binding. RNA 10(10):1563–1571. doi:10.1261/rna.7920904
115. Micklem DRAJGSSJD (2000) Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. EMBO J 19(6):1366–1377
116. Hough RF, Lingam AT, Bass BL (1999) Caenorhabditis elegans mRNAs that encode a protein similar to ADARs derive from an operon containing six genes. Nucleic Acids Res 27(17):3424–3432. doi:10.1093/nar/27.17.3424
117. Wang Z, Hartman E, Roy K, Chanfreau G, Feigon J (2011) Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. Structure 19(7):999–1010. doi:10.1016/j.str.2011.03.022
118. Connolly CM, Dearth AT, Braun RE (2005) Disruption of murine Tenr results in teratospermia and male infertility. Dev Biol 278(1):13–21. doi:10.1016/j.ydbio.2004.10.009
119. Melcher T, Maas S, Herb A, Sprengel R, Higuchi M, Seeburg PH (1996) RED2, a brain-specific member of the RNA-specific adenosine deaminase family. J Biol Chem 271(50):31795–31798. doi:10.1074/jbc.271.50.31795

120. Chen CX, Cho DS, Wang Q, Lai F, Carter KC, Nishikura K (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA-binding domains. RNA 6(5):755–767
121. Kim VN, Han J, Siomi MC (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10(2):126–139. doi:10.1038/nrm2632
122. Nemlich Y, Greenberg E, Ortenberg R, Besser MJ, Barshack I, Jacob-Hirsch J, Jacoby E, Eyal E, Rivkin L, Prieto VG, Chakravarti N, Duncan LM, Kallenberg DM, Galun E, Bennett DC, Amariglio N, Bar-Eli M, Schachter J, Rechavi G, Markel G (2013) MicroRNA-mediated loss of ADAR1 in metastatic melanoma promotes tumor growth. J Clin Invest 123(6):2703–2718. doi:10.1172/jci62980
123. Heraud-Farlow JE, Kiebler MA (2014) The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. Trends Neurosci. doi:10.1016/j.tins.2014.05.009
124. Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G (1993) The protein product of the fragile-X gene, Fmr1, has characteristics of an RNA-binding protein. Cell 74(2):291–298. doi:10.1016/0092-8674(93)90420-U
125. Ascano M Jr, Mukherjee N, Bandaru P, Miller JB, Nusbaum JD, Corcoran DL, Langlois C, Munschauer M, Dewell S, Hafner M, Williams Z, Ohler U, Tuschl T (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature 492(7429):382–386. doi:10.1038/nature11737
126. De Boulle K, Verkerk AJMH, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van Den Bos F, de Graaff E, Oostra BA, Willems PJ (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nat Genet 3(1):31–35
127. Siomi H, Choi MY, Siomi MC, Nussbaum RL, Dreyfuss G (1994) Essential role for Kh domains in RNA-binding—impaired RNA-binding by a mutation in the Kh domain of Fmr1 that causes fragile-X syndrome. Cell 77(1):33–39. doi:10.1016/0092-8674(94)90232-1
128. De Rubeis S, Fernández E, Buzzi A, Di Marino D, Bagni C (2012) Molecular and cellular aspects of mental retardation in the fragile X syndrome: from gene mutation/s to spine dysmorphogenesis. In: Kreutz MR, Sala C (eds) Synaptic plasticity, vol 970, Advances in experimental medicine and biology. Springer, Vienna, pp 517–551. doi:10.1007/978-3-7091-0932-8_23
129. Moritz M, Paulovich AG, Tsay YF, Woolford JL (1990) Depletion of yeast ribosomal proteins L16 or rp59 disrupts ribosome assembly. J Cell Biol 111(6):2261–2274. doi:10.1083/jcb.111.6.2261
130. Ebert BL, Pretz J, Bosco J, Chang CY, Tamayo P, Galili N, Raza A, Root DE, Attar E, Ellis SR, Golub TR (2008) Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. Nature 451(7176):335–339. doi:10.1038/nature06494
131. Hartner JC, Walkley CR, Lu J, Orkin SH (2009) ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. Nat Immunol 10(1):109–115. doi:10.1038/ni.1680

Chapter 9
Splicing Factor Mutations in Cancer

Rafael Bejar

Abstract Many cancers demonstrate aberrant splicing patterns that contribute to their development and progression. Recently, recurrent somatic mutations of genes encoding core subunits of the spliceosome have been identified in several different cancer types. These mutations are most common in hematologic malignancies like the myelodysplastic syndromes (MDS), acute myeloid leukemia, and chronic lymphocytic leukemia, but also in occur in several solid tumors at lower frequency. The most frequent mutations occur in *SF3B1*, *U2AF1*, *SRSF2*, and *ZRSR2* and are largely exclusive of each other. Mutations in *SF3B1*, *U2AF1*, and *SRSF2* acquire heterozygous missense mutations in specific codons, resembling oncogenes. *ZRSR2* mutations include clear loss-of-function variants, a pattern more common to tumor suppressor genes. These splicing factors are associated with distinct clinical phenotypes and patterns of mutation in different malignancies. Mutations have both diagnostic and prognostic relevance. Splicing factor mutations appear to affect only a minority of transcripts which show little overlap by mutation type. How differences in splicing caused by somatic mutations of spliceosome subunits lead to oncogenesis is not clear and may involve different targets in each disease type. However, cells with mutated splicing machinery may be particularly vulnerable to further disruption of the spliceosome suggesting a novel strategy for the targeted therapy of cancers.

Keywords Cancer • Oncogenesis • Splicing • Myelodysplastic syndromes • SF3B1 • SRSF2 • ZRSR2 • U2AF1 • Stem cell • Acute myeloid leukemia

R. Bejar, M.D., Ph.D. (✉)
Division of Hematology and Oncology, UC San Diego Moores Cancer Center, La Jolla, CA, USA
e-mail: rabejar@ucsd.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_9

1 Introduction

Alternative mRNA splicing is a well-established means of diversifying the coding potential of the 26,000+ human genes in to hundreds of thousands possible protein products [1]. Differences in isoform utilization have been tied to lineage specificity and the dynamic behavior of cells in various contexts [2]. In the case of cancer, dysregulated splicing of critical genes can promote and maintain the transformed cell state. Examples include alternative splicing of anti-apoptotic gene transcripts *BCL2* and *BCL-X*, *CD44*, and *GSK3B* [3, 4, 5, 6].

Substantial evidence implicates differential expression of RNA-binding proteins and post-translational modifications of splicing factors as mechanisms that regulate variable splicing in normal and cancer cells [7, 8]. However, it was unrecognized until recently that somatic mutations in the splicing machinery itself are common drivers of cancer in various cell types. The oncogenic molecular mechanisms engaged by these mutations remain poorly understood. The frequency and type of splicing factor mutations differs among tumor types as does the stage at which these mutations are acquired demonstrating substantial tissue specificity. The specific mutations identified have interesting properties that hint at their pathogenic function, such as mutual exclusivity, heterozygosity, and conservative amino acid substitutions. Understanding how these mutations lead to cancer may identify novel mechanisms of transformation and suggest new therapeutic targets [9]. Even without a comprehensive understanding of how mutated splicing factor act as oncogenes, their presence carries important clinical implications that vary across tumor types and disease contexts.

This chapter will describe the discovery of acquired splicing factor mutations in cancer and review the most frequently altered genes in this class along with their suspected mechanism of action and clinical relevance.

2 Discovery of Splicing Factor Mutations in Hematologic Malignancies

Mutations of the splicing machinery were first identified in tumor material from patients with myelodysplastic syndromes (MDS) [10–12]. MDS are a collection of clonal hematopoietic stem cell neoplasms associated with low blood cell counts, dysmorphic maturation of differentiating cells in the bone marrow, and an elevated risk of transformation into acute myeloid leukemia. Patients with MDS are clinically heterogeneous, spanning a spectrum of more indolent lower risk disease that can persist for years, to cases with higher risk disease and a life expectancy measured in weeks to months [13]. Studies on the molecular basis of MDS demonstrate that these disorders are as varied genetically as they are clinically. Over 50 recurrently mutated genes have been identified in MDS tumor material in a multitude of combinations with most patients harboring only a few mutations [14–16]. While no

single gene is mutated in the majority of MDS cases, a typical mutation of a splicing factor can be found in nearly 70 % of cases, making MDS the tumor type most commonly associated with mutations in this class of genes. Mutations in one of four genes (*SF3B1*, *U2AF1*, *SRSF2*, *and ZRSR2*) account for nearly two-thirds of these cases.

2.1 SF3B1

Splicing factor 3B1 (*SF3B1*) is an integral member of the U2 small ribonucleoprotein complex responsible for branch site recognition near the 3′ end of in pre-messenger RNA. Whole exome sequencing of patients with MDS identified recurrent *SF3B1* mutations in about a quarter of cases. The majority)of these patients had a common subtype of MDS known as refractory anemia with ring sideroblasts (RARS). The RARS subtype of MDS is characterized by mitochondrial deposits of precipitated iron in ring pattern around the nucleus of developing erythrocytes (red blood cell precursors). When *SF3B1* mutations are identified in other MDS subtypes, ring sideroblasts are often present in these cases as well, strongly linking *SF3B1* mutations to this particular morphologic feature. Subsequently, *SF3B1* mutations were identified in nearly 20 % of patients with CLL, a hematologic malignancy largely unrelated to MDS [17, 18].

The pattern of somatic mutation in *SF3B1* is not random. The majority of mutations are missense substitutions at codon 700 with lysine replaced by glutamic acid (K700E). Less frequently, recurrent missense mutations at a small number of other hotspots are identified. Overall, these tend to be fairly conservative substitutions predicted to have less impact on protein function than random missense mutations might. No truncating nonsense mutations or frameshift insertions or deletions are found. Mutations of *SF3B1* are always heterozygous to a wild type allele and are expressed at the mRNA level suggesting that they are translated into a mutant protein. Heterozygous missense mutations at specific codons and a lack of truncating lesions are a signature of oncogenes that experience a change or gain of function when mutated in cancer. In SF3B1, these mutational hotspots are concentrated in the middle of 4 contiguous HEAT domains of which the longer isoform has a total of 22 in its distal half (Fig. 9.1) [17]. Mutations are far from the portion of the protein involved in branch site recognition, suggesting that they might alter interactions with other subunits instead of affecting the RNA-binding properties of the protein directly. Alternatively, hotspot mutations in oncogenes can result in dominant negative activity where the mutant protein suppresses the function of the unmutated subunit. However, this does not appear to be the case with oncogenic SF3B1 mutations.

Murine studies demonstrate that total loss of Sf3b1 is lethal during embryogenesis, but heterozygous deletion results in only a mild phenotype characterized by subtle skeletal abnormalities and minimal changes in hematologic parameters. Subsequent studies of heterozygous *Sf3b1* deletion in the adult hematopoietic com-

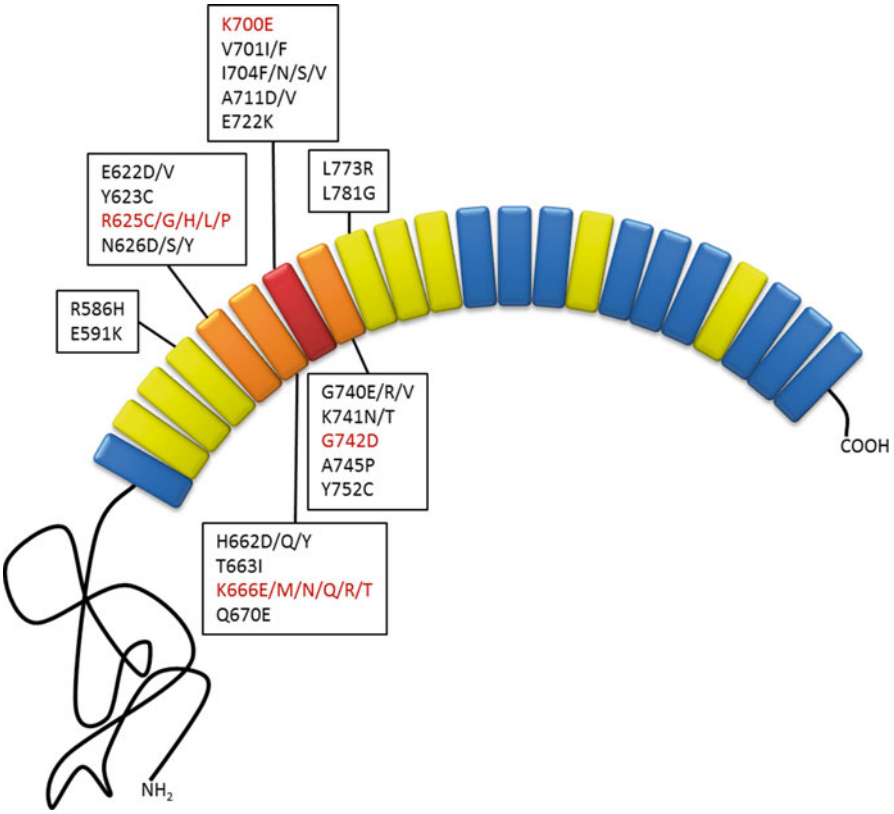


Fig. 9.1 Distribution of somatic mutations in SF3B1. Rectangles represent repetitive HEAT domains colored coded by the frequency with which they contain recurrent somatic mutations; *red*-highest, *orange*-high, *yellow*-low, *blue*-rare or none. Missense mutations in the most frequently mutated HEAT domains are listed. Those shown in red highlight the most frequently mutated codon in each HEAT domain. Mutation data from COSMIC was current as of 1/2015 and includes mutations reported in two or more samples. The figure is adapted from Bonnal et al. *Nature reviews. Drug discovery*. Nov 2012 [19]

partment have shown mixed results with one study identifying rare ring sideroblasts and another finding no increase in ring sideroblasts and a decrease in hematopoietic stem cells. Both studies reported only mild hematologic changes and no transformation to a clonal disorder such as myeloid leukemia or CLL [20–22]. Animal models expressing the hotspot mutations of *SF3B1* seen in patients have not been published to date, but studies of patient material have begun to shed light on how these lesions can affect splicing and expression patterns.

Gene expression profiles obtained from MDS and CLL patients with and without *SF3B1* mutations show some differences in mitochondria related gene signatures [10, 23] and possibly DNA damage responses [24]. However, exon utilization patterns are not strikingly different and the few genes with differential isoform expres-

sion do not always appear relevant to MDS or oncogenesis. RNA-sequencing (RNAseq) studies in a variety of tissue types have found that the predominant splicing abnormality in *SF3B1* mutant cells is the utilization of a cryptic splice acceptor just upstream of the canonical 3′ splice junction, often resulting in an out-of-frame splicing event [25].

In uveal melanoma, a solid tumor characterized by recurrent *SF3B1* mutations, several alternatively spliced candidate genes were identified in mutant cases. Examination of these alternative isoforms in MDS and CLL cases found that they were enriched in tumors with *SF3B1* mutations, suggesting a common signature associated with these lesions. How alternative isoforms created by *SF3B1* mutations lead to a clonal advantage and how they impact disease development remains unknown.

Differences in mutation patterns between MDS, CLL, and uveal melanoma demonstrate that there is some tissue specificity to how *SF3B1* mutations act (Table 9.1). For example, uveal melanomas predominantly acquire mutations at codon 625 and almost never develop the K700E substitution that accounts for half of the *SF3B1* mutations in CLL and MDS [26]. In MDS, mutations of *SF3B1* and several other splicing factors appear to be early events as they are typically found in the dominant tumor clone. In CLL, where the same codons are recurrently mutated at rates similar to MDS, *SF3B1* mutations are often subclonal indicating later acquisition, often at the time of relapsed disease. The clinical implications of identical *SF3B1* mutations are different in these two tumor types as well. In CLL, *SF3B1* mutations are associated with resistance to chemotherapy and shorter overall survival whereas in MDS, the same mutations predict a lower rate of transformation to AML and greater likelihood of having indolent disease [14, 15, 27, 28].

2.2 U2AF1

The U2-complex auxiliary factor 1 gene (*U2AF1*) encodes a 35-kDa subunit of the U2-spliceosome responsible for recognition of the terminal 3′ AG dinucleotide in pre-mRNA introns. The encoded protein has four major domains including two zinc finger regions, an arginine-serine domain, and a U2AF homology domain with which it heterodimerizes with its larger 65-kDa partner, U2AF2. These form a complex with additional factors to bind at the 3′ end of introns, recognizing the terminal splice acceptor site and its preceding degenerate polypyrimidine tract. U2AF1 interacts directly with several splicing factors known to develop somatic mutations in cancer including U2AF2, SRSF2, and SF1 although these abnormalities are mutually exclusive and have variable rates of mutation in different tumor types.

U2AF1 is most frequently mutated in MDS)with mutations found in approximately 10 % of cases. Mutations occur primarily at two hotspots, codons S34 and Q157, where several different missense substitutions can arise [12, 15]. These codons are in adjacent RNA-binding zinc-finger domains suggesting that they might alter recognition of splice sites. Rarer insertions and deletions near codon 157 that maintain the reading frame have also been described. As with *SF3B1*, mutations of

Table 9.1 Splicing factor mutations, clinical associations, and effects on splicing

Splicing factor	Cancer type	Incidence	Mutation types	Clinical associations	Major effect on splicing
SF3B1	MDS and MDS/MPN	25 %	Heterozygous missense at specific codons, K700E most common	Acquired early, ring sidero blasts, favorable prognosis	3′ splicing at cryptic acceptor site upstream on canonical site
	CLL	15 %	″″	Resistance, poor prognosis	″″
	Breast cancer	1–2 %	″″	More common in ER+ papillary (16 %) and mucinous (6 %) subtypes	″″
	Pancreatic	1–2 %	″″		″″
	Uveal melanoma	15 %	Codons, R625H and R625C most common	Favorable prognosis	″″
U2AF1	MDS	10 %	Heterozygous missense at specific codons, at S34 or Q157 most common		
	Lung		ĐĐ		
SRSF2	MDS	15 %	Heterozygous missense or in-frame involving codon P95	Poor prognosis	Small changes in the inclusion or exclusion of cassette exons
	CMML	40–50 %	″″	Poor prognosis, exclusive of EZH2 mutations	″″

ZRSR2	MDS	5 %	Out-of-frame insertion/ deletion, premature stop codon, splice site, and missense distributed along the gene	Most often in male patients given location on X-chromosome	Retained introns associated with U12-dependent splicing events
	T-ALL	4 %	″″	Thymic cvariant	
	BPDCN	~75 %	″″	Strong male predominance	
	Solid tumors	1 %	″″	Strong male predominance	
U2AF2	MDS	Out-of-frame insertion/ deletion, premature stop codon, splice site, and missense distributed along the gene			
SF3A1	MDS	″″			
PRPF8	MDS	″″		Poor prognosis	
SF1	MDS	″″	″″		

U2AF1 in MDS and other cancers are typically heterozygous to a wildtype allele and do not include truncating nonsense or frameshift mutations. This suggests an oncogenic gain-of-function phenotype.

In contrast to *SF3B1*, mutations of *U2AF1* are rarely associated with RARS and ring sideroblasts in MDS. *U2AF1* mutations are also not observed in CLL or uveal melanoma which might be expected if these splicing factor mutations had the same pathogenic effects. In fact, there is evidence to suggest that the S34 and Q157 mutations of U2AF1 have distinct effects on splicing and disease phenotypes. First, only the S34 codon is mutated in lung cancers whereas the Q157 locus is not [29]. In MDS, Q157 substitutions account for 25–35 % of *U2AF1* mutations [14, 15, 30]. RNAseq studies of patient material, mouse models, and cells lines expressing mutant *U2AF1* show that alterations of splicing involve a small minority of expressed genes with the most common event being changes in cassette exon utilization [25, 31, 32]. The genes affected by the presence of an S34 mutation show little overlap with those affected by the presence of a Q157 mutation when compared in isogenic cell lines. Analysis of the affected splice junctions reveals that mutations of each codon affect the recognition of different aspect of the 3′ splice region by U2AF1. The two most common S34 mutations, S34F and S34Y, lead to increased exon inclusion when an A or C is present at the −3 position just upstream of the AG dinucleotide [31, 33]. Exons that show increased skipping in the presence of *U2AF1* S34 mutations most often have a T at this position. In contrast, exon utilization changes in the presence of Q157 mutations show no difference in base composition at their −3 position. Instead, differentially included exons are more likely to have a G or a T at the +1 position (first base in the exon) whereas differentially excluded exons are more likely to have an A or a C as this locus. This suggests that *U2AF1* mutations directly affect RNA-binding leading to changes in exon inclusion and that the different mutations likely impact different exons [31, 34].

2.3 SRSF2

The serine/arginine-rich splicing factor 2 gene (*SRSF2*) is another frequently mutated component of the spliceosome involved in processing of the 3′ splice acceptor site. The SRSF2 subunit contains an RNA recognition motif as well as an RS domain rich in arginine and serine residues. It is involved in many RNA-related processes including spliceosome assembly of the U1 snRNP to the 5′ splice site, U2 snRNP binding at the branch point, and mRNA stabilization. SRSF2 can bind splicing enhancer sequences through which it is known to regulate alternative isoform expression [35, 36]. These splicing enhancer sites are often in exons close to the 3′ intron-exon boundary and typically contain a GGNG or CCNG sequence motif when recognized by SRSF2 [35].

As with *SF3B1* and *U2AF1*, somatic mutations of *SRSF2* are nearly always heterozygous missense mutations at a specific codon—specifically P95 for *SRSF2*. Several different amino acids may be substituted for proline at this position, including

histidine, leucine, and arginine. Insertions or deletions that maintain the reading frame but disrupt the proline at codon 95 are also observed in rare cases. In hematologic malignancies, *SRSF2* mutations are found most often in MDS and in chronic myelomonocytic leukemia (CMML) with incidences of 15 % and nearly 50 % respectively [12]. This splicing factor is not commonly mutated in CLL or other lymphoid neoplasms, nor have frequent recurrent mutations been identified in solid tumors.

Clinically, *SRSF2* mutations in myeloid disorders have been associated with shorter overall survival or an increased risk of AML, but these links are not always independent of other established risk factors [37–39]. Mutations of *SRSF2* are typically exclusive of mutations in *SF3B1*, *U2AF1*, and other splicing factors despite differences in clinical phenotypes. For example, MDS or CMML patients with *SRSF2* mutations are much less likely to have ring sideroblasts and tend to have increased levels of monocytes. *SRSF2* mutations are also exclusive of mutations in *EZH2*, an H3K27 histone methyltransferase that is the catalytic subunit of the protein repressive complex 2. This mutual exclusivity may, in part, reflect how SRSF2 mutations promote myeloid oncogenesis.

Studies of *SRSF2* P95 mutations find only small differences in gene expression and splice isoform utilization. This suggests that these oncogenic mutations do not cause a significant loss of protein function. Changes in cassette exon inclusion and exclusion are the most common events observed by RNAseq in models of missense SRSF2 P95 mutation [40]. These involve a small minority of expressed genes and most often, only slightly change the ratio between alternative isoforms. One of the frequently affected targets is the *EZH2* pre-mRNA. In the presence of *SRSF2* mutation, a cryptic exon is frequently included in the mature mRNA causing a change of reading frame [41]. This is associated with lower levels of the translated EZH2 protein, suggesting a convergent pathogenic mechanism and potentially explaining why somatic mutations of *SRSF2* and *EZH2* are exclusive of each other in patients with MDS [39].

A mouse model of *SRSF2* P95H expression in the mature hematopoietic compartment confirms that this lesion does not cause a total loss of function. In contrast to mice with *SRSF2* deletion in adult hematopoietic cells, which develop a hypocellular bone marrow and no dysplasia, the P95H mutant animals have normal marrow cellularity and marked myeloid dysplasia mimicking the human disease state. The P95H mutant mice also have an increased proportion of early stem/progenitor cells. RNAseq analysis demonstrates that preferentially excluded cassette exons in P95H mutant animals and model cell lines are enriched for GGNG at the exon splicing enhancer site. Preferentially included cassette exons are instead enriched for the CCNG motif [42].

2.4 ZRSR2

The zinc finger (CCCH type), RNA-binding motif and serine/arginine rich 2 gene (*ZRSR2*) is another SR-rich splicing factor frequently mutated in myeloid malignancies. The encoded protein is a component of the U2 auxiliary factor heterodimer and

participates in the recognition of the 3′ splice acceptor site. Its somatic mutation pattern is very different from that of *SF3B1*, *U2AF1*, and *SRSF2*. First, *ZRSR2* resides on the X-chromosome, meaning men only carry a single copy. Second, *ZRSR2* mutations are often out-of-frame insertions or deletions, splice-site mutations, or nonsense mutations predicted to prematurely truncate the full length protein. Missense mutations are also frequent, but are scattered throughout the length of the gene and not concentrated in specific regions or codons. This pattern of mutation is consistent with a loss-of-function. Therefore, it is not surprising that mutations occur more frequently in male patients in whom a single *ZRSR2* mutation would leave no remaining wildtype allele.

ZRSR2 is mutated in about 5 % of patients with MDS and in a similar fraction of those with T-cell acute lymphoblastic leukemia [11, 15, 43]. These mutations appear to be the predominant genetic lesion in patients with plasmacytoid dendritic cell neoplasms, a very rare lymphoid disorder characterized by strong male predominance [44]. *ZRSR2* mutations have also been found at lower frequency in a variety of other tumor types. Recent studies suggest that ZRSR2 is a critical element of the U12 minor spliceosome [45]. Absence of ZRSR2 had no effects on U2 complex dependent splicing events, but demonstrated marked aberrant splicing of U12-depedent events. Retained introns were the most commonly identified event. Which of the genes altered by *ZRSR2* mutation are most likely to cause transforming effects or disease features remains unknown.

3 Additional Splicing Factors

Several other splicing factors are known to be recurrently mutated in a variety of cancer types, albeit more rarely. The majority of these less frequently mutated genes appear to acquire truncating or other abnormalities predicted to cause a loss of function. These genes include *U2AF2*, *SF3A1*, *LUC7L2*, *SF1*, *PRPF40B* and *PRPF8* among several others [11, 14, 15, 46]. Given their relative scarcity, less is known about molecular mechanisms or phenotype specific associations for these genes.

4 Splicing Inhibitors

As a ubiquitous mechanism for the regulation of gene expression, mRNA splicing is required for the survival of essentially every human cell type, including cancerous ones. Animal studies have repeatedly demonstrated that complete loss of major splicing factors is not tolerated even when restricted to adult tissue types. This observation hints at the possibility that neoplastic cells with acquired abnormalities of their splicing machinery may have a selective sensitivity to further disruption of splicing [19]. Based on the observation that many cancers establish a program of altered splicing, several splicing factor inhibitors have been developed for human

use [47]. Some have even been tested in clinical trials although not specifically in patients with splicing factor mutations. The synthetic pladienolide derivative, E7107, is an inhibitor of the U2-spliceosome targeting its SF3B1 activity. This drug has demonstrated anti-tumor activity *in vitro* an *in vivo* cancer models. E7107 was first given to patients in phase I trial of advanced solid tumors to determine its safety [48]. There were limiting side-effects at doses capable of affecting splicing as measured in the peripheral blood. The most troubling of these was optic neuritis in several patients. While this effect appeared reversible, its cause is not understood. It is not clear if this is a class effect or simply specific to E7107.

Studies examining the vulnerability of cells with splicing factor mutations to splicing antagonists have not consistently identified a selective sensitivity associated with these mutations. For example, the spliceosome inhibitors, FD-895 and pladienolide B, cause rapid apoptosis in CLL cells from patients compared to normal B-cells [49]. However, this effect is independent of the cells' *SF3B1* mutations status. Since aberrant splicing in cancer can be caused by mechanisms other that splicing factor mutations, many cancer cells may have an inherent susceptibility to splicing inhibitors that is more universal and not necessarily mutation specific [50].

5 Summary

Acquired mutations of splicing factors are recurrent events in a variety of malignancies, particularly myeloid neoplasms such as the myelodysplastic syndromes. The three most frequently mutated splicing factors, *SF3B1*, *U2AF1*, and *SRSF2* share several similarities that suggest they might have a common mechanism of action. All three genes acquire heterozygous missense mutations at specific codons that are largely exclusive of each other. And, all three encode components of U2 spliceosome responsible for 3′ splice acceptor site recognition and excision. However, their effects on splicing and alternative isoform expression show little overlap. *SF3B1* mutations appear to alter 3′ splice site recognition, leading to utilization of cryptic upstream sites in a subset of genes. *U2AF1* and *SRSF2* mutations most often affect the inclusion or exclusion of cassette exons but show little overlap in the genes involved. Even the splicing events altered by the two major *U2AF1* mutations contain distinct recognition motifs and impact different genes. *ZRSR2* mutations, which are also exclusive of those in *SF3B1*, *U2AF1*, and *SRSF2*, appear to affect introns removed by the U12 minor splicing pathway while leaving U2-mediated events unaffected. Finally, there are several tumor types in which one splicing factor is mutated frequently, but the other two are not. For example *SF3B1* is mutated in CLL, breast cancer, and uveal melanoma, but *U2AF1* and *SRSF2* are not. In uveal melanoma, the *SF3B1* mutation most common to MDS, CLL, and breast cancer (K700E) is almost never seen and instead, the R625 substitutions predominate. These tissue type specific patterns of mutation would not be expected if all of these splicing factor genes affected a common set of transcripts to drive oncogenesis.

The alternative explanation is that splicing factor mutations regulate different subsets of target genes which may vary by the gene involved, the mutation type, and the cellular context. Their mutual exclusivity may be mediated by intolerance for compound abnormalities of the splicing machinery. If true, this would suggest a selective toxicity to pharmacogenic disruption of splicing. So far, early studies have shown that splicing inhibitors may have anti-neoplastic activity, but it is not necessarily associated with mutations of splicing factors.

The recent discovery of acquired splicing factor mutations in cancer has uncovered a novel mechanism by which neoplastic cells alter their genetic program to promote clonal selection. It has provided insight into the mechanisms responsible for oncogenesis and disease prevention. Most importantly, we may leverage this understanding to provide better diagnostic, prognostic, and therapeutic tools for the care and treatment of patients with malignancies.

References

1. La Cognata V, Iemmolo R, D'Agata V et al (2014) Increasing the coding potential of genomes through alternative splicing: the case of PARK2 gene. Curr Genomics 15:203–216
2. Dutertre M, Sanchez G, Barbier J et al (2011) The emerging role of pre-messenger RNA splicing in stress responses: sending alternative messages and silent messengers. RNA Biol 8:740–747
3. Ghigna C, Valacca C, Biamonti G. (2008) Alternative splicing and tumor progression. Current genomics 9(8):556–70. PubMed PMID: 19516963; PubMed Central PMCID: PMC2694562
4. Wang L, Zuo B, Xu D, Ren Z, Zhang H, Li X, et al. (2012) Alternative splicing of the porcine glycogen synthase kinase 3beta (GSK-3beta) gene with differential expression patterns and regulatory functions. PLoS One. 7(7):e40250. PubMed PMID:22792253; PubMed Central PMCID:PMC3391277
5. Pal S, Gupta R, Davuluri RV (2012) Alternative transcription and alternative splicing in cancer. Pharmacol Ther 136:283–294
6. Shkreta L, Bell B, Revil T et al (2013) Cancer-associated perturbations in alternative pre-messenger RNA splicing. Cancer Treat Res 158:41–94
7. Prochazka L, Tesarik R, Turanek J (2014) Regulation of alternative splicing of CD44 in cancer. Cell Signal 26:2234–2239
8. Brown RL, Reinke LM, Damerow MS et al (2011) CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. J Clin Invest 121:1064–1074
9. Yoshida K, Ogawa S (2014) Splicing factor mutations and cancer. Wiley Interdiscip Rev RNA 5:445–459
10. Papaemmanuil E, Cazzola M, Boultwood J et al (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 365:1384–1395
11. Yoshida K, Sanada M, Shiraishi Y et al (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478:64–69
12. Makishima H, Visconte V, Sakaguchi H et al (2012) Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 119:3203–3210
13. Bejar R (2014) Clinical and genetic predictors of prognosis in myelodysplastic syndromes. Haematologica 99:956–964
14. Papaemmanuil E, Gerstung M, Malcovati L et al (2013) Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122:3616–3627

15. Haferlach T, Nagata Y, Grossmann V et al (2014) Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 28:241–247
16. Bejar R, Stevenson K, Abdel-Wahab O et al (2011) Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med 364:2496–2506
17. Quesada V, Conde L, Villamor N et al (2012) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44:47–52
18. Wang L, Lawrence MS, Wan Y et al (2011) SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med 365:2497–2506
19. Bonnal S, Vigevani L, Valcarcel J (2012) The spliceosome as a target of novel antitumour drugs. Nat Rev Drug Discov 11:847–859
20. Matsunawa M, Yamamoto R, Sanada M et al (2014) Haploinsufficiency of Sf3b1 leads to compromised stem cell function but not to myelodysplasia. Leukemia 28:1844–1850
21. Visconte V, Rogers HJ, Singh J et al (2012) SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. Blood 120:3173–3186
22. Visconte V, Tabarroki A, Zhang L et al (2014) Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk Myelodysplastic syndromes with ring sideroblasts. J Hematol Oncol 7:89
23. Dolatshad H, Pellagatti A, Fernandez-Mercado M et al (2015) Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia 29:1092–1103
24. Te Raa GD, Derks IA, Navrkalova V et al (2014) The impact of SF3B1 mutations in CLL on the DNA-damage response., Leukemia
25. Shiozawa Y, Sato-Otsubo S, Gallì A et al (2014) Comprehensive analysis of aberrant RNA splicing in myelodysplastic syndromes. Blood 124:826
26. Field MG, Harbour JW (2014) Recent developments in prognostic and predictive testing in uveal melanoma. Curr Opin Ophthalmol 25:234–239
27. Malcovati L, Papaemmanuil E, Bowen DT et al (2011) Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. Blood 118:6239–6246
28. Baliakas P, Hadzidimitriou A, Sutton LA et al (2014) Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia 29(2):329–336
29. Imielinski M, Berger AH, Hammerman PS et al (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150:1107–1120
30. Bejar R, Stevenson KE, Caughey BA et al (2012) Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J Clin Oncol 30:3376–3382
31. Ilagan JO, Ramakrishnan A, Hayes B et al (2015) U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res 25:14–26
32. Shirai CL, Ley JN, White BS et al (2014) Mutant U2AF1 expression alters hematopoiesis and Pre-mRNA splicing in transgenic mice. Blood 124:827
33. Brooks AN, Choi PS, de Waal L et al (2014) A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS One 9, e87361
34. Shao C, Yang B, Wu T et al (2014) Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. Nat Struct Mol Biol 21:997–1005
35. Moon H, Cho S, Loh TJ et al (2014) SRSF2 promotes splicing and transcription of exon 11 included isoform in Ron proto-oncogene. Biochim Biophys Acta 1839:1132–1140
36. Pandit S, Zhou Y, Shiue L et al (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol Cell 50:223–235
37. Mian SA, Smith AE, Kulasekararaj AG et al (2013) Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica 98:1058–1066

38. Patnaik MM, Lasho TL, Finke CM et al (2013) Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. Am J Hematol 88:201–206
39. Meggendorfer M, Roller A, Haferlach T et al (2012) SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood 120:3080–3088
40. Zhang J, Lieu YK, Ali AM et al (2015) Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proc Natl Acad Sci U S A 112:E4726–E4734
41. Kim E, Ilagan JO, Liang Y et al (2015) SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell 27:617–630
42. Kim E, Ilagan JO, Lee S et al (2014) SRSF2 mutations impair hematopoietic differentiation by altering exonic splicing enhancer preference. Blood 124:824
43. Neumann M, Vosberg S, Schlee C et al (2015) Mutational spectrum of adult T-ALL. Oncotarget 6:2754–2766
44. Kim SS, Stevenson KE, Yoda A et al (2013) Loss-of-function mutations in the splicing factor ZRSR2 Are common in Blastic plasmacytoid dendritic cell neoplasm and have male predominance. Blood 122:741
45. Madan V, Kanojia D, Li J et al (2015) Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nat Commun 6:6042
46. Kurtovic-Kozaric A, Przychodzen B, Singh J et al (2015) PRPF8 defects cause missplicing in myeloid malignancies. Leukemia 29:126–136
47. Dehm SM (2013) Test-firing ammunition for spliceosome inhibition in cancer. Clin Cancer Res 19:6064–6066
48. Eskens FA, Ramos FJ, Burger H et al (2013) Phase I pharmacokinetic and pharmacodynamic study of the first-in-class spliceosome inhibitor E7107 in patients with advanced solid tumors. Clin Cancer Res 19:6296–6304
49. Kashyap MK, Kumar D, Villa R et al (2015) Targeting the spliceosome in chronic lymphocytic leukemia with the macrolides FD-895 and pladienolide-B. Haematologica 100:945–954
50. Hsu TY, Simon LM, Neill NJ et al (2015) The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature 525:384–388

Chapter 10
Regulation of Tissue-Specific Alternative Splicing: *C. elegans* as a Model System

Xicotencatl Gracida*, Adam D. Norris*, and John A. Calarco

Abstract Alternative pre-mRNA splicing serves as an elegant mechanism for generating transcriptomic and proteomic diversity between cell and tissue types. In this chapter, we highlight key concepts and outstanding goals in studies of tissue and cell-specific alternative splicing. We place particular emphasis on the use of *C. elegans* as a tractable model organism for *in vivo* studies of alternative splicing between tissues and also at single cell resolution. We describe our current understanding of tissue and cell-specific regulation in the animal, and emerging techniques that will allow for future mechanistic studies as well as systems level investigations of spatio-temporal splicing under laboratory conditions and in response to environmental stimuli.

Keywords Alternative splicing • RNA processing • *C. elegans* • Tissue-specific regulation

1 Introduction

1.1 Importance of Alternative Splicing in Generating Diversity, Specialization and Regulation

In engineering, the complexity of a system often correlates with the number of components and the interactions between them. The same principle generally applies to biological systems, where cellular and molecular complexity seems to arise from the number of different macromolecular building blocks and their interactions. As such, mechanisms to diversify these building blocks have evolved in all living organisms. For example, the path to creating proteins and modulating their

*These authors contributed equally to this work.

X. Gracida • A.D. Norris • J.A. Calarco (✉)
FAS Center for Systems Biology, Harvard University,
52 Oxford Street, Cambridge, MA 02138, USA
e-mail: jcalarco@fas.harvard.edu

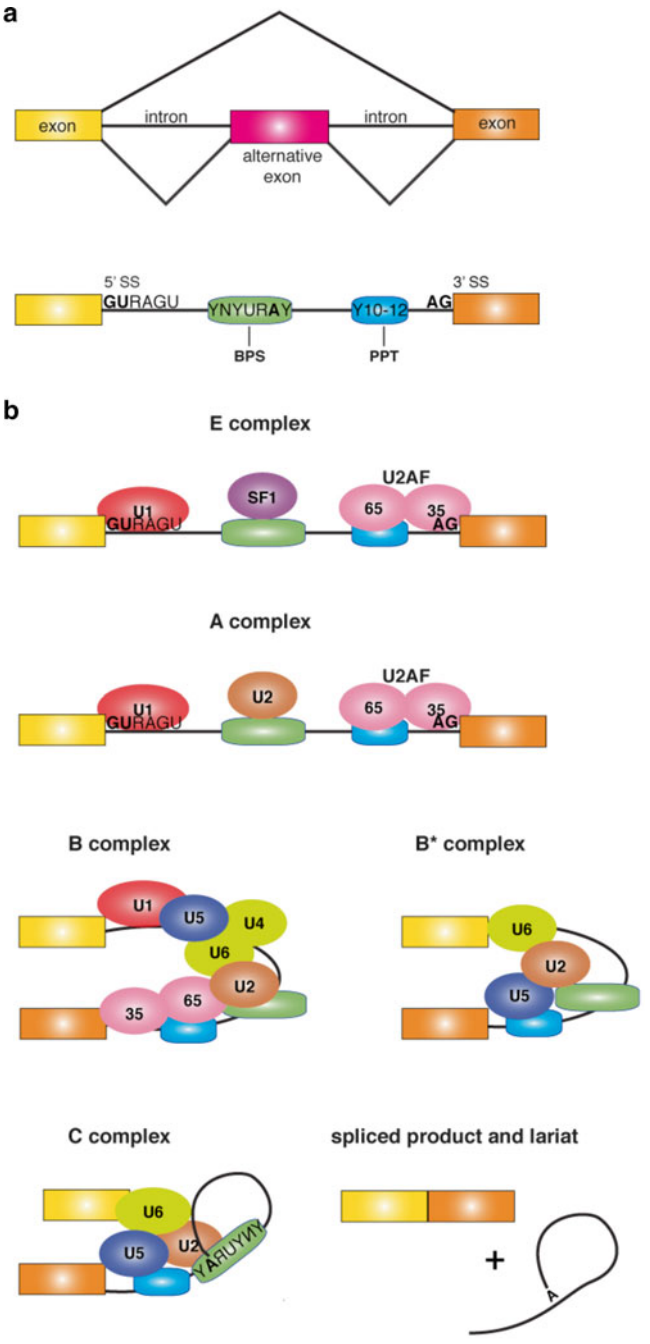
G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_10

abundance and/or activity involves several layers of transcriptional, post-transcriptional, translational and post-translational regulation, all of which have likely evolved to create increased potential for diversification. The co- and post-transcriptional regulatory toolkit alone is broad, including distinct types of regulatory proteins as well as short and long regulatory RNAs such as micro RNAs and long non-coding RNAs [1–3]. One key co- and post-transcriptional mechanism that serves to expand the cellular repertoire of expressed proteins isoforms is through alternative mRNA splicing: the process in which different mRNAs are generated from a single precursor RNA (pre-mRNA) [4, 5].

In most eukaryotes, in order to generate a mature mRNA with an uninterrupted open reading frame (ORF), non-coding intronic sequences are removed from nascent pre-mRNA and the remaining protein-coding exonic sequences are ligated together. This process is called mRNA splicing and it is an essential processing step for the proper expression of most genes in metazoans [4, 5]. Depending on the frequency in which a given exon is present in an mRNA, it can be classified as *constitutively spliced* or *alternatively spliced* (Fig. 10.1a). Constitutively spliced exons are always present in a given mRNA, whereas alternatively spliced exons can be selectively included or excluded at specific stages of development or in a tissue-specific manner [6]. A remarkable example of how alternative splicing can generate molecular diversity is the case of the *Drosophila* immunoglobulin superfamily gene Down syndrome cell adhesion molecule 1 (Dscam1). Through the combinatorial use of four sets of alternative exons (containing 12, 48, 33, and 2 variants each) alternative splicing of this gene could potentially generate up to 38,000 distinct Dscam1 isoforms [7]. Several reports have demonstrated that distinct Dscam1 variants are required for proper establishment of connectivity in the fly nervous system [8, 9].

What features determine whether an exon will be alternatively spliced, and how is this process regulated in a tissue-specific manner? To answer some of these questions we will provide a brief overview of the mechanism of RNA splicing and the enzymatic machinery that regulates this process. We will then discuss insights on how tissue-specific alternative splicing is regulated, and how this can help to generate distinct cellular subtypes with specialized functions. For the remainder of the review, we will focus on the nematode *Caenorhabditis elegans*, an emerging model for studying tissue specific alternative splicing, and we present technologies that are being developed and currently applied in this system.

Fig. 10.1 Pre-mRNA splicing and spliceosome assembly. (**a**) Example of alternative splicing involving cassette-exon inclusion or skipping. Below are depicted *cis*-acting element sequences involved in pre-mRNA splicing. 5′ and 3′ splice sites (SS), Branch Point Sequence (BPS) and Polypyrimidine Tract (PPT). For consensus sequences: N is any nucleotide, R is a purine, and Y is a pyrimidine. (**b**) Spliceosome assembly across intron. U1, U2, U5, U6 are small nuclear RNPs; U2AF is U2 auxiliary factor complex composed of the 35 and 65 subunits; SF1 is splicing factor 1. Complex A represents the pre-spliceosome. Complex B and B* the pre-catalytic and activated spliceosome, respectively. Complex C represents the catalytic step 1 spliceosome. See main text for additional explanation


a
exon
intron
alternative exon
intron
exon
5' SS
GURAGU
YNYURAY
Y10-12
AG
3' SS
BPS
PPT
b
E complex
U1
GURAGU
SF1
U2AF
65
35
AG
A complex
U1
GURAGU
U2
U2AF
65
35
AG
B complex
U1
U5
U4
U6
U2
35
65
B* complex
U6
U2
U5
C complex
U6
U5
U2
YARUYNY
spliced product and lariat
+
A

2 Mechanisms of Splicing

The enzymatic complex that removes introns and joins exons from pre-mRNAs is known as the spliceosome. This complex is composed of proteins and small nuclear RNAs that interact with conserved core signals (*cis*-elements) present in the pre-mRNA, that include the 5′ and 3′ splice sites, the branch point sequence (BPS), and the polypyrimidine tract (PPT) (Fig. 10.1a) [10, 11]. The main components of the spliceosome are five Uridine-rich small nuclear RNAs called U1, U2, U4, U5 and U6, which along with a variable number of proteins form distinct small nuclear ribonucleoproteins (snRNPs) [10].

The components of the spliceosome assemble and release sequentially until it becomes catalytically active. The first step in the spliceosome assembly involves base-pairing of U1 snRNP to the intron 5′ splice-site (Fig. 10.1b). Then the splicing factor 1 (SF1) protein binds to the branch point sequence, and the U2 auxiliary factor heterodimer, composed of U2AF65 and U2AF35, binds to the polypyrimidine tract and 3′ splice-site, respectively, forming the E complex [10, 11]. In the next step of assembly the U2 snRNP replaces SF1 and base-pairs with the branch point sequence to form the pre-spliceosomal A complex (Fig. 10.1b). Subsequently, the U4/U5 and U6 snRNPs are recruited as a pre-assembled complex. At this point, the resulting assembled B complex is still catalytically inactive. Activation of the complex B does not occur until the snRNAs are structurally rearranged and form novel base-pairing interactions, and U1 and U4 are removed forming a catalytically active C complex [10, 11]. Through a transesterification reaction, the first catalytic step generates a free 5′ exon and an intron lariat-3′ exon (Fig. 10.1b). After additional rearrangements, a reactive region in the 5′ exon attacks the 3′ splice site via a second transesterification reaction that ultimately ligates the exons and excises the lariat intron. Recently, single molecule fluorescence studies on spliceosome catalysis have indicated that each of these dynamic steps can be reversible [12]. Moreover, because the complex lacks a preformed active center, its sequential assembly and extensive structural rearrangements make the spliceosome prone to extensive regulation and to have a relatively high flexibility in target recognition.

3 Regulation of Splice-Site Recognition: Auxiliary *cis*-Elements and Regulatory Factors

Most of the RNA-RNA interactions in the splicing process are not strong, but they are enhanced or stabilized by different *trans*-acting proteins that interact with flanking pre-mRNA *cis*-regulatory sequences. Initial recognition of the 5′ and 3′ target splice-sites in the pre-mRNA occurs through base-pairing with snRNPs. The strength of this base-pairing is influenced by the actual consensus sequence of the splice-site, which in most higher eukaryotes are relatively degenerate [11, 13]. The *cis*-elements that regulate splice-site recognition can be classified as intronic or

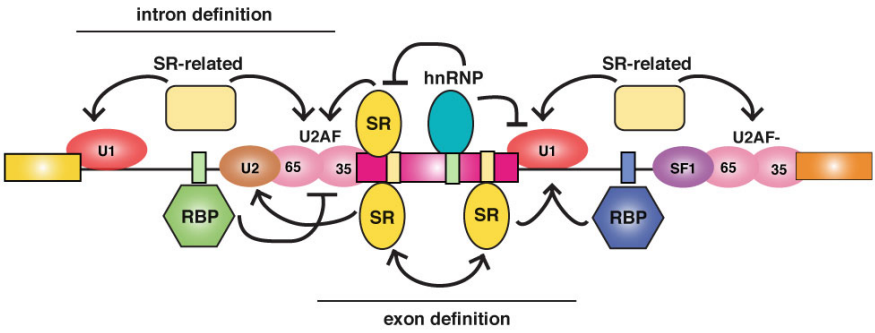


Fig. 10.2 Combinatorial regulation of alternative splicing. Cartoon shows two models of exon splicing: intron and exon definition through interaction of *cis*-acting elements (*squares*) and *trans*-acting elements. RNA-binding proteins (RBP), Ser-Arg protein family (SR), heterogeneous nuclear ribonucleoprotein (hnRNP). U1 and U2 are small nuclear RNPs; U2AF is U2 auxiliary factor complex composed of the 35 and 65 subunits; SF1 is splicing factor 1. See main text for additional explanation

exonic, and as splicing enhancers or silencers, depending on their position and their effect on splice-site usage (Fig. 10.2) [14, 15]. Most of these *cis*-elements are thought to interact with *trans*-acting proteins that activate or repress recruitment of the splicing machinery.

Two main broad classes of *trans*-factors regulate splicing decisions: the Ser-Arg protein family (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Fig. 10.2). These two classes of proteins were initially proposed to mediate antagonistic effects: SR proteins would primarily activate splicing by binding exonic enhancers, whereas hnRNPs would repress splicing by binding silencer elements [16, 17]. The combination of these factors (enhancers vs. silencers) with degenerate splice-site sequences would ultimately influence whether or not an exon is recognized for inclusion in an mRNA [11]. Although SR proteins and hnRNP proteins are still generally associated with enhancement and repression of splicing, respectively, several exceptions have been identified (for example: [18, 19]). Depending on *cis*-information in the surrounding context SR proteins and hnRNPs can act either as enhancers or silencers. Additionally, depending on the actual exon, the same splicing regulatory protein can promote either inclusion or exclusion. For example, hnRNP L regulates splicing in a context-dependent fashion [20], mediating repression when it is bound to an exon flanked by strong splice-sites, and mediating inclusion when bound to an exon flanked by weak splice-sites. How does a single protein mediate both effects is proposed to arise not only from the contextual information but ultimately on how the factor impinges on rate-limiting steps in spliceosome assembly [20].

Aside from hnRNP and SR proteins, it is believed that a suite of other RNA-binding proteins also play key roles in modulating splicing decisions through the recognition of *cis*-elements (Fig. 10.2). Some examples of these factors will be discussed further in the next section. Finally, there are additional features known to

influence whether an exon is constitutively or alternatively spliced, such as the elongation rate of RNA polymerase II [21], chromatin structure [22–25], sizes of exons and introns, and presence of secondary structures in the pre-mRNA [26, 27]. For detailed descriptions of these interconnections, several recent excellent reviews can be consulted [4, 28].

4 Splicing Factors and Tissue Specific Alternative Splicing

In contrast to the large number of characterized sequence-specific DNA binding proteins that regulate transcription, there are fewer characterized RNA-binding proteins that regulate splicing in a sequence and gene-specific manner. Based on current knowledge, two main mechanisms have been suggested for how alternative splicing is regulated in a spatiotemporal manner. First, differences in the activity (i.e. posttranslational modifications) and/or concentration of broadly expressed splicing regulators contributes to tissue- or stage-specific splicing outcomes (for example: [29–31]). Second, the action of tissue-restricted splicing factors enhance or repress splicing events in the corresponding tissue of interest (for examples, see [32–34]).

The brain is probably the best-characterized organ in regard to tissue- specific expression of splicing factors. Several conserved proteins have been found to have a restricted expression in the brain or in specific sub-regions, including members of the Nova, Fox, TIA, Quaking, CELF, and Hu/ELAV families of RNA-binding proteins. For a more inclusive list of factors and details on characterization of these factors we refer to several detailed reviews [6, 35, 36]. Here we will focus on one example to highlight how tissue-restricted expression of splicing factors can modulate tissue-specific exon networks.

The polypyrimidine-tract binding proteins (PTBP1 and PTBP2) are paralogous RNA-binding proteins involved in the regulation of cell-specific alternative splicing. PTBP1 is broadly expressed, and it is thought to repress inclusion of neuron-specific exons in non-neuronal tissues. PTBP1 also inhibits the inclusion of an exon in PTBP2 transcripts, producing a transcript that contains a premature termination codon that is degraded by the NMD pathway [37, 38]. However, at the onset of neuronal differentiation PTBP1 expression is repressed by the microRNA mir-124, stimulating PTBP2 expression [39]. Recent reports suggest that embryonic PTBP2 controls neuronal maturation by repressing splicing of “adult-specific” exons, because in the absence of PTBP2 neural cells fail to mature and eventually die [40, 41].

It remains an interesting question as to how these two paralogs with highly similar RNA-binding domains can mediate partially non-overlapping regulation. Initial experiments studying the PTBP-regulated neural-specific src N1 exon suggested that the PTBP paralogs have different inhibitory strengths [42]. A recent study also suggests that sequential reduction of PTBP1 and then PTBP2 expression define distinct stages of neural differentiation and splicing outcomes [43]. Moreover, additional cell-specific regulators have been shown to further influence

splicing events controlled by PTBP proteins. One of these factors is the neuronal-enriched SR-related protein of 100 kDa (nSR100/SRRM4). nSR100 can act to oppose PTBP-mediated exon skipping to facilitate exon inclusion of a network of neuron-specific alternative splicing events, including a set of recently discovered microexons that are misregulated in the brains of individuals with autism spectrum disorders [44–46].

Mechanistically, nSR100 has recently been found to promote exon inclusion in the nervous system by binding to UGC-containing intronic enhancers proximal to weak 3′ splice-sites and by recruiting early components of the spliceosome [46]. These interactions are thought to act in a dominant positive manner, counteracting the repressive effects of PTBP1. Interestingly, in non-neuronal tissues these UGC-containing elements serve to actively weaken 3′ splice-site recognition, leading to exon exclusion [46]. Taken together, these studies exemplify how combinations of factors cooperate to regulate networks of alternatively spliced isoforms that contribute to the identity and function of a specific tissue.

5 Importance of Alternative Splicing in Defining and Functionalizing Tissues

The dynamic nature of the splicing machinery makes it prone to extensive regulation that is sensitive to cellular environments. As a result, alternative splicing can generate tissue- or cell-type specific signatures of expressed isoforms that contribute to the specialization or function of a tissue. Accordingly, recent high-throughput studies in several organisms have confirmed that alternative mRNA isoforms are frequently differentially-expressed between tissues or associated with particular developmental or cell states (for example, see: [47–54]).

In one recent example, embryonic stem (ES) cells and other undifferentiated pluripotent cells express a distinct collection of spliced isoforms relative to differentiated cells [55]. Using high-throughput RNA-sequencing and computational analyses, a recent report found that the muscleblind-like RNA-binding proteins (MBNL1/2) repress a collection of alternative splicing events that promote ES cell state [55]. Among these ES cell-defining isoforms is a splice variant of the forkhead family transcription factor FOXP1, which stimulates expression of other pluripotency-associated genes [56].

The development of a tissue can also be accompanied by coordinated changes in alternative splicing. During heart development and remodeling, several genes were found to be temporally coordinated in their relative proportions of alternative isoforms. These differences in isoform abundance are regulated by concomitant developmental changes in the concentrations of the MBNL and CELF splicing factors, which bind regulatory sequences in the coordinated genes [33]. Furthermore, CELF concentration level is regulated by a microRNA, highlighting the cross-regulation of alternative splicing and microRNA regulation [57]. Additionally, the Epithelial Splicing Regulatory Proteins 1 and 2 (ESRP1 and 2) were identified as tissue-

specific regulators of a network of alternative splicing events important for the epithelial cell state [34]. It was later found that loss of these RNA-binding proteins led to splicing changes and morphological transformations similar to the epithelial to mesenchymal transition found both in development and in cancer [58].

A recurrent theme is that alternative splicing is often interconnected with other layers of gene expression to aid in the functionalization of a tissue [4]. One example of this occurs in the nervous system where the splicing factor nSR100/SRRM4 mentioned above and the RE1-silencing transcription factor (REST) show cross-regulation [59]. nSR100 mediates inclusion of an exon in REST transcripts, and the resulting REST isoform is truncated and has a less repressive transcriptional activity, allowing the expression of genes that are otherwise repressed in non-neuronal tissues where full-length REST is expressed. One of the genes repressed by REST in non-neuronal tissues is nSR100 thus creating a double negative feedback loop [59]. Overall, these examples show that several developmental and differentiation programs employ alternative splicing for the regulation of cellular transitions that define or help in the functionalization of a tissue.

Apart from generating molecular diversity by increasing the proteomic repertoire, alternative splicing is also an important mechanism for regulating expression levels of mRNA and protein. One mechanism is through increasing the regulatory plasticity of an mRNA through alternative usage of 5′ and 3′ un-translated regions (UTRs), which may influence mRNA localization, stability or translation efficiency [60, 61]. Another mechanism occurs through coupling alternative splicing with the nonsense-mediated mRNA decay (NMD) pathway, which targets mRNAs with premature termination codons for degradation. One function of the NMD pathway is to clear potentially detrimental mRNAs generated by errors in transcription or splicing [62]. However, through the inclusion or exclusion of alternatively spliced exons that alter the reading frame of an mRNA, coupling of alternative splicing to NMD serves to play a role for regulating abundance of isoforms through a mechanism that is independent of transcriptional regulation [63]. Often, transcripts subject to alternative splicing-NMD coupled regulation encode RNA-binding proteins, providing an elegant mechanism for auto- and cross-regulation that likely play important roles in establishing and maintaining robust isoform networks during cellular differentiation (for examples see [64–70]).

6 Constitutive and Alternative Splicing in *C. elegans*

The basic principles of constitutive and alternative splicing are shared throughout metazoans. *C. elegans* genes, like those in vertebrates, are intron rich. Human and worm transcripts are similar with regards to the number of exons per gene and the average exon size, but they diverge in that human introns are roughly ten times larger than *C. elegans* introns (average intron length of ~1000 nucleotides in humans versus ~100 nucleotides in *C. elegans*) [71, 72]. *C. elegans* 5′ and 3′ splice-site consensus sequences are essentially the same as those in vertebrates, except that *C.*

elegans introns have a highly-conserved polyU tract immediately upstream of the 3′ splice-site [71]. Interestingly, many *C. elegans* transcripts undergo *trans*-splicing, where capped splice leader sequences transcribed from distinct loci can be spliced to the 5′ end of mRNA transcripts. This review will not cover *trans*-splicing, but see [73] for a more detailed review. Importantly, human transcripts exhibit a significantly greater degree of alternative splicing than worm transcripts. Current estimates suggest that ~25 % of genes in *C. elegans* undergo alternative splicing [53], while in humans that number is upwards of 95 % [15, 51]. Indeed, it has been demonstrated that in metazoans the degree of alternative splicing scales with rough measures of organismal complexity [74], although these correlations are still somewhat controversial. However, *C. elegans* does exhibit a rich diversity of alternative splicing, including all common regimes of alternative splicing observed in vertebrates (*e.g.*, cassette exons, mutually-exclusive exons, alternative 5′ and 3′ splice-sites, etc.), and regulation of alternative splicing occurs by similar mechanisms [72]. Therefore the simplicity and experimental manipulations available in *C. elegans* make it an ideal model organism in which to study alternative splicing and its regulation. In particular, its transparency, invariant cell lineage, and powerful genetic tools make it an excellent tool for elucidating mechanisms of alternative splicing at the cell- and tissue-specific level.

7 Monitoring Alternative Splicing in *C. elegans*

Many examples of alternatively spliced genes in *C. elegans* have come to light fortuitously in the course of cloning and characterizing individual genes (for example: [75–79]). Later, analysis of expressed sequence tags (ESTs) allowed for a more systematic search for alternative splicing in *C. elegans*, detecting evidence for hundreds of alternatively spliced genes [80]. More recently, high-throughput techniques have enabled more comprehensive detection of alternative splicing across the entire transcriptome. First, splicing-sensitive microarrays were developed for use in *C. elegans*, where they were used to monitor alternative splicing of hundreds of exons undergoing changes throughout development and under genetic perturbations [53, 64, 65, 81]. Next-generation sequencing has afforded even deeper detection of alternative splicing, providing evidence for thousands of alternative splicing events in *C. elegans*, and showing that hundreds of these are developmentally-regulated [49, 53]. Such studies have updated the estimated percentage of *C. elegans* genes undergoing alternative splicing to ~25 %, and this number may continue to grow as transcriptomes are analyzed from worms under additional conditions and at greater depth. Deep sequencing has also been recently used to identify target transcripts of factors regulating alternative splicing, detecting hundreds of splicing events under the control of RNA-binding proteins in *C. elegans* [82, 83].

8 Tissue Specific Alternative Splicing in *C. elegans*

8.1 Observing Tissue-Specific Alternative Splicing

While the approaches discussed above are powerful for detecting alternative splicing across the transcriptome, they are limited to giving information about the whole organism, and unable to provide information about alternative splicing between tissues or cell types. Methods used in mammals and *Drosophila* for profiling the transcriptomes of dissected tissues or cell lines originating from different tissues [47, 48, 50, 51, 54] are not presently feasible in *C. elegans* due to the technical challenges of dissecting tissues and establishing cell lines in the worm. However, recent advances in *C. elegans* cell dissociation followed by FACS sorting, in which tissue-specific promoters are used to mark and sort tissues of interest (i.e. [84–86]), followed by RNA sequencing, hold promise for future studies measuring such tissue-specific transcriptome profiling in the worm. An alternative strategy is to express an epitope-tagged mRNA-binding protein, such as Poly-A Binding Protein, in the tissue of interest, then co-immunoprecipitate the mRNA-binding protein with its associated mRNAs. The immunoprecipitated mRNA can then be profiled by microarray analysis. This approach, known as mRNA tagging has been used to identify mRNAs expressed in specific tissues [85, 87, 88] and holds promise for identifying tissue-specific isoforms as well.

On the other hand, the worm is also ideally suited for observing individual alternative splicing events with tissue specificity *in vivo*, due to its transparency and reproducible cell lineage. One particularly fruitful strategy has involved multi-color mini gene fluorescent reporter systems, in which an alternative exon (or exons) plus the upstream and downstream introns and constitutive exons are placed upstream of fluorescent protein open reading frames. The classic version of the system relies on tandem green and red fluorescent proteins (GFP and RFP) encoded in alternate reading frames [89, 90] (Fig. 10.3a). The reporter is then engineered such that expression of one isoform drives productive expression of GFP while expression of the other isoform shifts the reading frame and drives expression of RFP. These reporters and other versions (Fig. 10.3b, c) provide a two-color readout of alternative splicing as controlled by the endogenous splicing machinery in the tissues in which the reporter is expressed. This system was originally designed for use in cell culture, but has been adapted successfully for use *in vivo* in the worm, where such reporters can be used not only to visualize alternative splicing within tissues or throughout development, but also to genetically identify factors controlling such alternative splicing [83, 91].

The first such reporter used in *C. elegans* demonstrated that the ortholog of the Fibroblast Growth Factor (FGF) receptor *egl-15* encodes a mutually-exclusive exon 5 in which 5A is selected in body-wall and vulval musculature, while 5B is selected in the hypodermis [91]. These results dovetail nicely with previous work showing that the 5A isoform is required for sex myoblast migration, while the 5B isoform is essential for viability [92]. Similar reporters were used to visualize isoforms of *unc-60*, *C.*

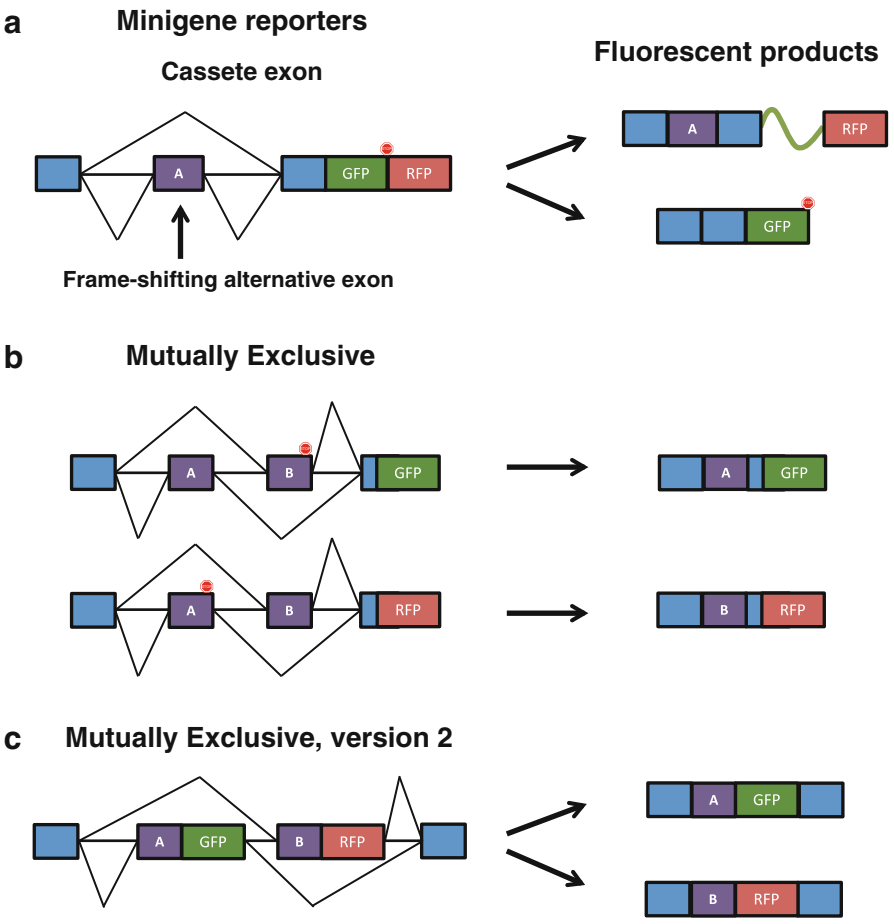


Fig. 10.3 Two-color splicing reporters used in *C. elegans*. (**a**) Reporter for cassette exons. A single transgene is produced in which GFP and RFP are in different reading frames, and the cassette exon encodes a +1 nt frameshift such that when excluded, GFP is produced in frame followed by a stop codon. When included, GFP is read out-of-frame without stop codons, and RFP is produced in frame. (**b**) Reporter for mutually exclusive exons. Two different transgenes are constructed. To one a stop codon is added to the end of exon B, and to the other a stop codon is added to exon A. Therefore GFP from construct one will only be produced when exon A is chosen and RFP from construct two will only be produced when exon B is chosen. (**c**) An alternative strategy for reporting mutually exclusive alternative splicing. A single construct is created in which GFP is fused to the 3′ end of exon A while RFP is fused to the 3′ end of exon B

elegans homolog of the actin-binding protein ADF/Cofilin, demonstrating that muscle tissue expresses the B isoform, while the A isoform is expressed in non-muscle tissues including the nervous system and intestine [93].

The above experiments monitored alternative splicing events that were already known by functional analyses to have tissue-specific isoform function, and the value in using the splicing reporters was to visually clarify expression of the

isoforms, and to use the reporters to genetically identify regulators of the splicing events. More recently, these reporters have also been used to provide new information on tissue-specific isoforms. For example, the *unc-32* gene encoding a subunit of the vacuolar-type H^+-ATPase had previously been suggested as a developmentally-regulated alternatively spliced gene [53], and one of its three exon 4 variants had been identified as neuronally-enriched [94], but not until the construction of three-color fluorescent splicing reporters was it clear that all three exon 4 isoforms are regulated tissue-specifically. It was demonstrated that exon 4A is predominantly expressed in the intestine, 4B in the nervous system and 4C in the pharynx [95]. Moreover, the *unc-32* transcript contains two mutually exclusive seventh exons, in which exon 7A is expressed predominantly in head neurons, while exon 7B is expressed in most other tissues [95]. These results highlight the power of multi-color fluorescent splicing reporter systems for describing new tissue-specific isoforms, even in genes undergoing extensive alternative splicing.

8.2 Assigning Function to Tissue-Specific Isoforms

Despite widespread observation of alternative splicing differences among mammalian tissues, it is difficult to gain insights regarding the functional relevance (if any) of individual tissue-specific isoforms. Studies in *C. elegans* have provided a platform for transgenic analysis of tissue-specific isoform function. For example, as mentioned above the *unc-32* gene encodes a subunit of the V-ATPase in which exon 4 is chosen in a mutually-exclusive manner from among 3 exons. One isoform is expressed in the nervous system, one in the intestine and one in the pharynx. Loss of the neuronal isoform leads to uncoordinated movement, whereas loss of the other isoforms leads to lethality in larval stages [94]. However, it was recently demonstrated that any of the other exon 4 isoforms can rescue for loss of exon 4B isoforms [95].

In another example, the Actin Depolymerizing Factor ADF/cofilin *unc-60*, discussed above, is alternatively spliced such that UNC-60A is expressed ubiquitously while UNC-60B is expressed in muscle. In concordance with these observations, disruption of UNC-60A caused embryonic lethality, while disruption of UNC-60B resulted in viable worms with disorganized actin filaments in muscle [96]. Although these expression patterns suggested distinct functional roles for each isoform, subsequent experiments demonstrated that when mis-expressed in muscle cells, UNC-60A could indeed compensate for the loss of UNC-60B [93, 97]. These examples highlight two interesting concepts. First, they demonstrate an additional value of using the worm to study isoform-specific function, in that the original isoform-specific phenotypes can be discovered due to genetic lesions that specifically reside in isoforms causing tissue-specific phenotypes. Second, these studies highlight the importance of assessing function of isoforms *in vivo*, and suggest that tissue-specific expression need not always imply unique function.

A final example lies in the two isoforms of the FGF receptor *egl-15*, which as described above, are differentially expressed between hypodermis (EGL-15B) and musculature (EGL-15A). Loss of EGL-15B results in lethality, while loss of EGL-15A results in specific defects in myoblast migration [92] . In this model alternative splicing event, however, EGL-15B was subsequently demonstrated to not rescue the myoblast migration defects resulting from loss of EGL-15A, indicating that these two isoforms possess unique functions [98]. Additionally, it was also found that EGL-15A plays an important role in maintaining axon positions via its expression in hypodermal cells [99]. Again, EGL-15B was not able to compensate for the role of EGL-15A in axon maintenance.

9 Cell-Specific Alternative Splicing Within Tissues

9.1 Observing Cell-Type Specific Alternative Splicing

The examples above demonstrate that fluorescent splicing reporters can be useful for observing tissue-specific alternative splicing, but such reporters are particularly powerful when applied in higher-resolution studies, visualizing alternative splicing between individual cell subtypes within tissues. This level of resolution has not been feasible in mammalian studies of alternative splicing, although recent reports have begun examining alternative splicing among different brain regions [100, 101]. On the other hand, invertebrate model systems have recently provided information about alternative splicing at the level of individual neurons [83, 102]. *C. elegans* is particularly well-suited for such studies at single-neuron resolution, with a reproducible cell lineage and transparent body. To look for cell-specific alternative splicing within the nervous system, we created two-color splicing reporters to monitor several different alternative splicing events in genes with known neuronal expression. Among the events we tested, roughly half exhibited neuron-specific alternative splicing where the splicing pattern in one population of neurons was different than the splicing pattern in another population of neurons [83]. Each event had a unique neuron-specific splicing profile, but each profile was completely reproducible between animals, suggesting these neuron-specific splicing events are highly regulated. Neuron-specific alternative splicing patterns included isoforms which were differentially regulated between inhibitory GABAergic motorneurons versus excitatory cholinergic motorneurons, isoforms which were differentially regulated between head neurons and pharyngeal neurons, and isoforms which were differentially regulated between oxygen sensing and mechanosensory neurons [83]. Such results can guide future functional tests for isoform-specific function tailored to the neuron types of interest.

Recent studies have provided additional examples of neuron-subtype specific isoform usage. For example, the neuronal-secreted Punctin *madd*-4 encodes a long isoform expressed only in cholinergic motor neurons and a short isoform expressed in both cholinergic and GABAergic motor neurons [103]. The proper expression of

these isoforms is essential for appropriate post-synaptic receptor expression, as mis-expressing the long isoform in GABAergic neurons causes recruitment of cholinergic receptors to GABAergic synapses, while disrupting the short isoform causes recruitment of GABA receptors to cholinergic synapses [103].

Taking the concept one step further, a new isoform of the sole *C. elegans* insulin receptor *daf-2* was recently discovered to be highly expressed in sensory neurons [104]. This isoform is specifically required for conditioned starvation avoidance learning, and it was shown that this variant (DAF-2C) is specifically re-localized from the cell bodies to the axons of chemosensory neurons following starvation, while other isoforms remain localized to the cell body [104]. Therefore, not only is this isoform expressed neuron subtype-specifically, it is also found localized dynamically in specific neuronal compartments. In this case, the chemosensory neuron-enriched isoform (but not other isoforms) becomes axonally localized following a starvation stimulus, and this splicing mediated re-localization is necessary and sufficient for specifying a *daf-2* that responds to the starvation stimulus. Additional isoforms of *daf-2*, when engineered to be localized to axons, were capable of restoring the conditioned starvation response, so the function of DAF-2C at least in this context is not unique, but rather it is the coupling of alternative splicing with re-localization of the protein that is important [104]. It will be interesting to learn whether other *daf-2* isoforms have specific functions in other cell types. More broadly it will be interesting to learn how widespread such isoform-specific function of cell-specific alternative splicing events is in *C. elegans*.

Moving forward, tissue- and cell type-specific transcriptome-wide analyses of wild type and splicing factor mutant strains will identify programs of coordinately regulated splicing events most likely to play functional roles in animal development and physiology. This approach has already proven effective using whole animal transcriptomes [64, 65, 81–83]. For example, analysis of alternative splicing events regulated by UNC-75 and/or EXC-7 led to the identification of a number of candidate alternative splicing events enriched in genes with roles in synaptic transmission. Further study of one such splicing event in the synatxin ortholog *unc-64* revealed non-redundant roles for the two isoforms in coordinating proper locomotory activity [83].

10 RNA-binding Proteins and *cis*-Elements Controlling Cell Specific Alternative Splicing in *C. elegans*

In this section we will present mechanistic details for how tissue-restricted alternative splicing in *C. elegans* can be achieved. The RNA-binding proteins and *cis*-elements that regulate these events in *C. elegans* (Table 10.1) are highly conserved, further validating the use of the system to understand mRNA splicing. In many cases the logic to understand tissue-specific regulation includes four general principles. First, one or more RNA-binding proteins are often required to bind in *trans* to the pre-mRNA. Second, the RNA-binding proteins generally

Table 10.1 Table of RNA-binding proteins in *C. elegans* known to regulate alternative splicing

Name/family	Binding domain	Binding motif	Regulated tissue	Target gene	regulated AS event	Known additional factors	Reference:
MEC-8/	RRM	U/AGCACA	Mechanosensory neurons and hypodermis	*unc-52*	Mutually exclusive	–	[127]
ASD-1/FOX-1	RRM	UGCAUG	Muscle, neurons	*egl-15*,, *unc-32*	Mutual exclusive	SUP-12	[91]
SUP-12	RRM	GUGUG	muscle	*egl-15*, *unc-60*	Mutual exclusive	ASD-1	[105]
ASD-2b/STAR	KH	CUAACUCUAAC	muscle	*let-2*	Mutually exclusive	–	[108]
		CUAAC	muscle	*unc-60*	Multiple exon skipping	SUP-12	[93]
HRP-2/hnRNP Q, R	RRM	UCUAUC	not determined	*unc-52*, *lin-10*	Cassette exon	–	[119]
UNC-75/CELF	RRM	UUGUUGUGUUGU	neurons	*unc-32*/ATPase V_0 subunit	Mutually exclusive	ASD-1, FOX-1	[95]
		U and G rich					
		UGUUGUG	GABAergic neurons	*unc-16*/JIP3 kinase	Cassette exon	EXC-7	[83, 127]
EXC-7/Hu/ELAV	RRM	UAAGUU	cholinergic neurons	*unc-16*	Cassette exon	UNC-75	[83, 127]

Table highlighting RNA-binding proteins found to regulate alternative splicing with identified *cis*-elements. Also listed are the type of RNA-binding domain, tissues undergoing regulation, key model target genes, type of splicing event regulated, known RNA-binding protein co-regulators, and references where *cis*-elements were identified. RRM stands for RNA recognition motif; KH stands for K-Homology domain

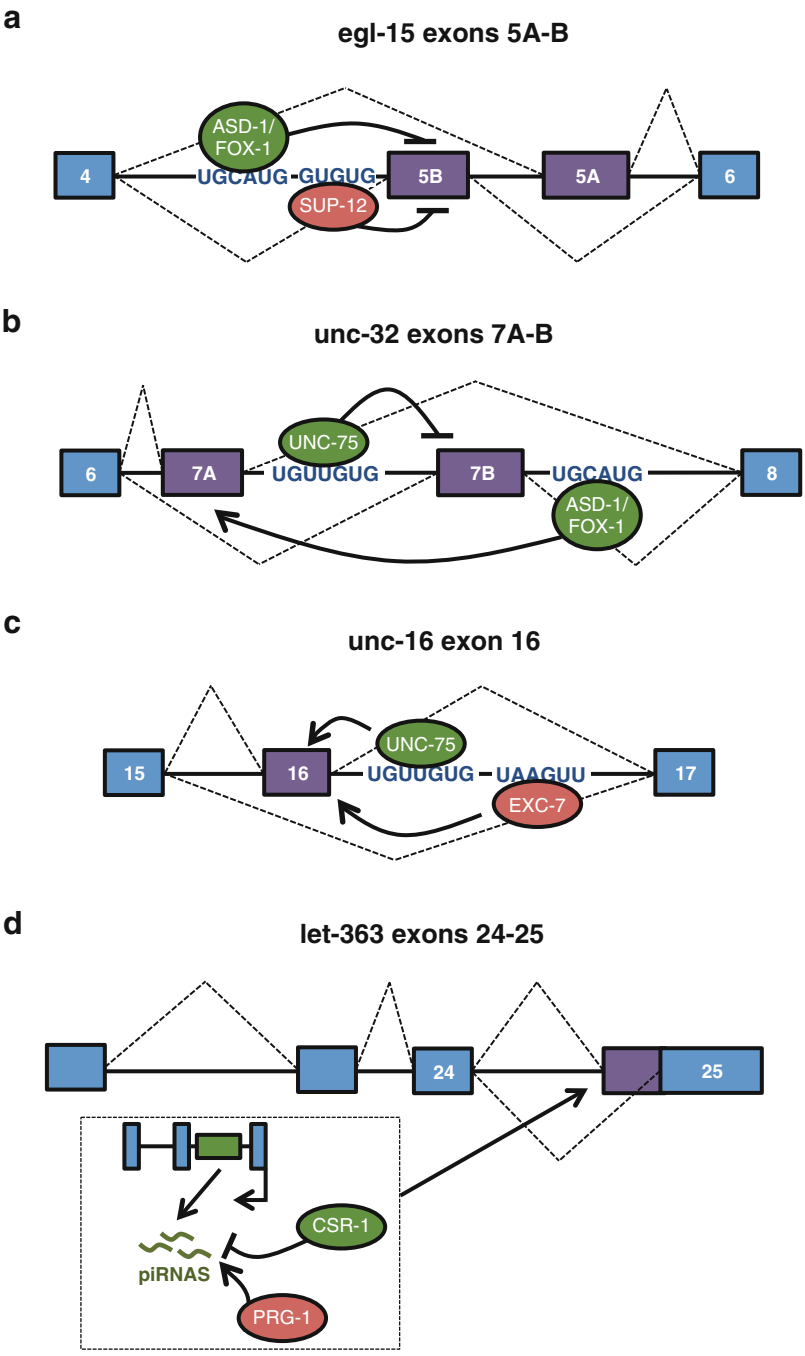
exhibit sequence-specific or structural preferences for *cis* elements in the pre-mRNA. Third, the relative position and context of these and other *cis*-regulatory elements in the pre-mRNA with respect to alternatively spliced exons matters. Finally, additional auxiliary proteins that synergize, antagonize, or potentiate the effect of the pre-mRNA bound RNA-binding protein on RNA splicing. *C. elegans* provides an attractive model system for studies of these principles *in vivo*.

As mentioned above, a breakthrough in the study of tissue-specific alternative splicing in *C. elegans* came from the use of fluorescent splicing reporters to visualize the output of splicing in different tissues [91]. In *C. elegans* the fibroblast growth factor receptor (FGFR) *egl-15* contains two mutually exclusive exons (5B and 5A, located upstream and downstream, respectively) that show tissue-specific expression patterns (Fig. 10.4a). Kuroyanagi *et al.* [91] used the fluorescent reporter system in a genetic screen to identify genes controlling the tissue-restricted splicing patterns of *egl-15* isoforms. ASD-1 (alternative splicing defective 1), an RNA-binding protein of the Fox-1 family, was identified to promote exon 5A inclusion in body wall muscle, which along with FOX-1, another family member, bind the highly conserved motif UGCAUG located downstream of intron 4. This binding results in repression of exon 5B thus allowing the use and inclusion of exon 5A in muscle [91].

Additional proteins were found to cooperatively regulate correct splicing of *egl-15*. SUP-12 is an RNA-binding protein that binds both ASD-1 and FOX-1 proteins and a GUGU stretch in *egl-15* intron 4 to cooperatively repress exon 5B in muscle [105] (Fig. 10.4a). In *sup-12* mutants, the muscle-specific inclusion of exon 5A is diminished or lost, highlighting the requirement of SUP-12 along with ASD-1 and FOX-1 to robustly repress exon 5B [105]. Recent structural data using ASD-1 and SUP-12 RNA-recognition motifs (RRM) in complex with target RNA sequence indicate that robust exon repression arises from both RRMs "sandwiching" a G base in intron 4 to form a tight ternary complex with a high repressive activity on exon 5B [106, 107]. Future studies will be needed to determine how the positioning of this complex actually represses exon 5B to give rise to a muscle-specific alternative splicing pattern. One possibility may be their binding position relative to the 3′ splice site. SUP-12 binds in intron 4, and the "G sandwiching", occurs around the place where branch point binding proteins would bind, and therefore an attractive hypothesis is that SUP-12 interferes with U2 snRNP recruitment when regulating this splicing event [105].

Further studies also found that ASD-1 and FOX-1 cooperatively regulate tissue-specific alternative splicing in neurons. Analogous to the need of a muscle-specific factor (SUP-12) for muscle-specific splicing, ASD-1 and FOX-1 in neurons require the neuron-specific RNA-binding protein of the CELF family UNC-75 [95]. As mentioned above, the H+-ATPase *unc-32* gene contains two sets of mutually

Fig. 10.4 Examples of combinatorial control of cell/tissue-specific alternative splicing. (**a**) ASD-1 or FOX-1 and SUP-12 control alternative splicing of the mutually exclusive exon #5 of *egl-15*. (**b**) ASD-1 or FOX-1 and UNC-75 control alternative splicing of mutually exclusive exon #7 of *unc-32*. (**c**) UNC-75 and EXC-7 control alternative splicing of *unc-16* cassette exon #16. (**d**) CSR-1 and PRG-1 control production of piRNAs in response to an intronic antisense transposable element, which affect the 5′ splice site selection of exon #25 of *let-363/TOR*


a
egl-15 exons 5A-B
ASD-1/
FOX-1
4
UGCAUG
GUGUG
5B
5A
6
SUP-12
b
unc-32 exons 7A-B
UNC-75
6
7A
UGUUGUG
7B
UGCAUG
8
ASD-1/
FOX-1
c
unc-16 exon 16
UNC-75
15
16
UGUUGUG
UAAGUU
17
EXC-7
d
let-363 exons 24-25
24
25
CSR-1
piRNAS
PRG-1

exclusive exons (4A,B, and C, and 7A and B) that are differentially expressed among tissues. Using fluorescent reporters monitoring these mutually exclusive *unc-32* exons, a forward genetic screen, and biochemical data, it was demonstrated that ASD-1 and FOX-1 bind a UGCAUG stretch in intron 7b, stimulating the selection of exon 7a in neurons [95] (Fig. 10.4b). In addition, exon 7a selection requires UNC-75, which binds at a conserved UUGUUGUGUUGU stretch in the intron 7a, further tipping the balance towards promoting selection of exon 7a and suppressing exon 7b usage [95] (Fig. 10.4b). UNC-75 was also found to play a key role in promoting inclusion of neuronal exon 4b in *unc-32* transcripts. It is unclear whether the same RNA-binding protein involved in splicing exon 4b also later plays a role in regulating exon 7a splicing. These results raise the interesting possibility that coordination of splicing at distinct sites in the same mRNA transcript may occur.

In addition to tissue-regulated exons, *C. elegans* extensively regulates alternative splicing during development [49, 53, 65]. One example of interest is the *let-2* pre-mRNA, containing two mutually exclusive exons 9 and 10, which undergo developmentally controlled alternative splicing in muscle cells [108]. In embryos and younger animals exon 9 is used, whereas in older animals exon 10 is used. ASD-2, a novel member of the signal transduction activators of RNA (STAR) family regulates this switch. A muscle-restricted isoform of ASD-2, *asd-2b*, accumulates in older animals where it binds an intronic stretch with strong similarity to the Quaking/QKI response element found in mammals (NACUAYY-N1-20-UAAY) [108, 109]. The authors put forward a model in which at early stages and in absence of ASD-2 a strong 5′ splice site in intron 9 versus a weaker 5′ splice site in intron 10 favors splicing of exon 9 to exon 11. However, later in development the binding of ASD-2 to the QKI response element-like stretch in the intron 10 favors the splicing of exon 10 to exon 11. Once this commitment has been made, the upstream exon 8 can then be spliced with either exon 9 or 10 [108]. These results suggest that the ordering of the splicing reaction need not occur in a linear contiguous manner, and may contribute to the underlying regulation of splice site choice, and have been conceptually found in mammals as well [110].

To further add to the regulatory complexity in *C. elegans*, distinct combinations of the same RNA-binding proteins can control separate splicing events. In a study of the splicing of *unc-60* transcripts, two RNA-binding proteins—ASD-2 and SUP-12 mentioned in examples above—were found to promote the expression of the UNC-60B isoform in muscle cells [93]. This muscle-specific isoform results from skipping of several exons (exons 2A-5A) downstream of exon 1 and splicing of exon 1 to a distal exon 2B. This pattern suggests that both SUP-12 and ASD-2 must suppress splice site pairing of exon 1 with exon 2A long enough for the 3′ splice site flanking exon 2B to be transcribed. In agreement with this mechanism, both factors were found to bind cooperatively to cis-elements located in intron 1 [93]. Similar to the regulation of *egl-15* alternative splicing, it is thought that ASD-2 and SUP-12 may sterically inhibit recruitment of U2 snRNP to the branch site via the U2 auxiliary factor complex (U2AF65/35) [93]. Taken together, these analyses demonstrate that the presence of distinct regulatory motifs combined with expression of multiple tissue specific factors enables combinatorial and context-dependent regulation of alternative splicing.

It appears that the regulatory principles mentioned above also apply to differential splicing patterns between individual neuronal subtypes. In a recent study, we analyzed the splicing patterns of several genes expressed in the *C. elegans* nervous system using two color reporters. Intriguingly, we found frequent differential regulation of splicing patterns between individual neuronal subtypes [83]. One such reporter monitored alternative splicing of exon 16 in *unc-16*, a gene encoding a protein localized to the axonal initial segment involved in regulating the selective transport of cargo molecules to axons [111, 112]. This alternative exon was found to be differentially spliced between cholinergic and GABAergic motor neurons in the ventral nerve cord, and a genetic screen identified two highly conserved RNA-binding proteins—UNC-75/CELF and EXC-7/Hu/ELAV—that control this neuron-specific splicing outcome by binding to *cis*-elements in the intron downstream of the alternative exon [83] (Fig. 10.4c). In the ventral nerve cord, UNC-75 is expressed in both cholinergic and GABAergic motor neurons, while EXC-7 expression is restricted to cholinergic neurons [113, 114]. Additional experiments suggest a model where the partially-overlapping expression patterns of these two factors govern the specificity of the *unc-16* exon 16 splicing pattern.

Several studies in mammalian cells have provided links between argonaute proteins involved in small RNA-mediated pathways, chromatin modifications, RNA polymerase II elongation, and regulation of alternative splicing [115, 116]. A recent study in *C. elegans* has now suggested that endogenous small RNAs may play a role in regulating alternative splicing of the Target of Rapamycin (TOR) gene *let-363* [117]. Using RNA-Seq, Barberan-Soler and colleagues identified several antisense transcripts enriched in sperm. One particular transcript (B0261.6) is transcribed on the antisense strand and within an intron of *let-363*, and contains a Helitron transposable element (Fig. 10.4d). Further experiments found that the presence of the transposable element in the antisense transcript leads to the production of endogenous small RNAs, the production of which are stimulated and repressed by argonaute proteins PRG-1 and CSR-1, respectively. The *let-363* locus contains an alternative splicing event that is nearby the Helitron transposable element and is differentially spliced between sperm and oocytes [117] (Fig. 10.4d). Interestingly, in *prg-1* and *csr-1* mutant animals, the *let-363* alternative splicing event is reciprocally affected, suggesting that these argonaute proteins antagonize each other in modulating small RNA levels that somehow influence splicing patterns. The effects on *let-363* splicing can also be achieved when exogenous double-stranded RNA targeting the transposon is introduced, in agreement with a mechanism where small RNA pathways are influencing splicing outcomes at this locus. Finally, animals exposed to this exogenous double-stranded RNA targeting the transposon were found to be fertile for many more generations than animals exposed to control double-stranded RNAs [117]. Collectively, this work suggests an interesting connection between small RNA pathways, tissue-specific splicing regulation, and heritable physiological traits. It will be interesting to see how widespread this mechanism is in *C. elegans* and other organisms.

In a different approach to identify *cis*-regulatory elements of alternative splicing in *C. elegans*, Kabat *et al*. used a bioinformatics approach to search for highly conserved regions in the flanking introns of 147 alternatively spliced cassette exons

[118]. Many pentamer and hexamer motifs were found enriched in either the upstream or downstream intron. Several identified motifs resemble known mammalian splicing regulatory sequences, highlighting again the conservation of splicing regulation in *C. elegans* [118]. Follow-up studies using RNA affinity chromatography and mass spectrometry analysis identified HRP-2, a *C. elegans* hnRNPQ/R homolog that binds UCUAUC motifs present in the *unc-52* gene. HRP-2 binds this intronic regulatory sequence flanking cassette exons where it can then promote inclusion of *unc-52* exons and also *lin-10* [119]. With the emergence of new datasets identifying a larger set of alternative exons and developmentally regulated splicing events, additional computational motif searches have been performed, identifying candidate *cis*-elements [53, 82, 83]. As future tissue-specific transcriptome data are generated, an additional layer of resolution should be achieved when searching for sequence elements that are important for splicing regulation in particular cell types.

Since RNA-binding proteins generally bind small motifs, it is often the case that a single splicing factor can regulate multiple mRNAs. One of the earliest reports in *C. elegans* of such a case was presented for the hnRNP F/H homologs *hrpf-1* and *sym-2*, which seem to control at least 31 developmentally regulated alternative spliced events [65]. With the advance of RNA-Seq based analysis of the transcriptomes of wild type and RNA-binding protein mutant animals, the discovery of events regulated by a single splicing factor has increased, identifying functionally coherent networks of even hundreds of events regulated by single or combinations of splicing factors [82, 83]. To determine whether events regulated by splicing factors are likely to be direct targets, combining information about the sequence specificities of RNA-binding proteins should be obtained through focused and/or genome-wide approaches. Two methods, systematic evolution of ligands by exponential enrichment (SELEX) and cross-linking and immunoprecipitation (CLIP) have paved the way to identify and map a big number of protein-RNA interaction motifs [120–122]. Both CLIP and SELEX have been successfully applied to C. elegans RNA-binding proteins (for examples see [123–125]), and should yield future insights. More recent *in vitro* techniques may offer a complementary approach to mapping protein-RNA interactions. The RNAcompete method was developed for systematic analysis of RNA-binding preferences and specificities [126, 127]. In this method, a cDNA library is cleaved off of a microarray and serves as a template to generate a pool of ~200 thousand RNAs (29–38 nucleotides) that are incubated *in vitro* with an RNA-binding protein of interest. After incubation, the bound RNA is then recovered and profiled. One of the limitations of this technique is that it does not integrate effects of secondary structure. However, the main advantage of this method is that it allows high-throughput identification of the relative preference of an RNA-binding protein for candidate *cis*-elements [126, 127].

A more recent and also high-throughput technique is RNA Bind-n-Seq [128], which is based on high-throughput SELEX and DNA Bind-n-Seq [129, 130]. Similar to these techniques, RNA Bind-n-Seq uses a one-step binding of a purified, tagged RNA-binding protein with a pool of random RNA oligonucleotides. The bound RNAs are then sequenced. Since RNA Bind-n-Seq uses different

concentrations of the same tested RNA-binding protein (i.e. low, medium, and high), the motif read enrichment at different concentrations is modeled to estimate dissociation constants [128]. The main advantage of RNA Bind-n-Seq is that it provides quantitative information about the affinity of an RNA-binding protein for its cognate motif. The information from all of the techniques described can be used to help predict networks of events regulated by existing and novel RNA-binding proteins and contribute to our understanding of the codes and regulatory logic governing splicing decisions.

11 Evolution of Tissue-specific Alternative Splicing in *C. elegans*

11.1 Evolutionary Characteristics of Alternative Splicing from Worm to Human

The dynamics of alternative splicing evolution appear to have proceeded significantly differently in nematodes than in other eukaryotic lineages. Comparisons between human and mouse or *Drosophila* and mosquito suggested that conservation of alternative exons between genomes are relatively lower than what is observed for constitutive exons [131, 132]. For instance, about 75 % of alternatively spliced exons in humans are also present in the mouse genome, and only 42 % of alternative exons in *Drosophila melanogaster* are also found in *Anopholes gambiae* [131, 132]. However, a similar comparison between *C. elegans* and *C. briggsae*—diverged by roughly the same distance as human and mouse—revealed that over 90 % of alternative exons in *C. elegans* are also found in the genome of *C. briggsae* [133]. More strikingly, when the analysis is restricted to "minor isoforms" (those comprising less than one third of the total abundance of transcripts for an individual gene), less than 27 % of this class of alternative exons is conserved from mouse to human, while 81 % are conserved between the two nematode species [133].

Such findings suggest that evolution of many alternative splicing events in *Caenorhabditis* nematodes may be more conservative than it is in other eukaryotic lineages, despite the fact that *Caenorhabditis* shows otherwise high rates of genomic evolution (*e.g.* sequence evolution, genome rearrangements, *etc.*) [133]. As discussed above, *Caenorhabditis* introns are much smaller than those found in vertebrates, and consequently are more densely populated in *cis*-regulatory information at the nucleotide level. It has been suggested that this may lead to introns less tolerant of deletions or insertions, which may in turn lead to restricted evolutionary space for alternative splicing relative to organisms such as vertebrates with significantly larger introns [65]. Another interpretation of these observations is that the evolution of alternative isoforms may have played a more important role in contributing to development and physiology of the *Caenorhabditis* genus rather than contributing to species-specific differences.

11.2 Evolution of Tissue-Specific Alternative Splicing

Studies in mammals have demonstrated that tissue-specific alternative splicing events are particularly highly conserved (for example see [134–136]), and studies in *Caenorhabditis* species have also shown a high degree of conservation at the sequence level and relative isoform abundance of alternative splicing events [53, 137, 138]. Conservation of tissue-specific alternative splicing in *Caenorhabditis* species has not directly been tested, but given the generally high conservation of alternative exons among *Caenorhabditis* it is reasonable to expect high levels of tissue-specific alternative splicing conservation. The emerging tissue-specific techniques discussed above should provide a powerful experimental platform for assessing conservation of tissue-specific alternative splicing in *Caenorhabditis* species, enabling studies at the intersection of evolutionary and developmental biology.

11.3 Conservation of Tissue-Specific Splicing Factors

Numerous important tissue-specific splicing factors discovered in mammals are conserved in *C. elegans*, with their tissue-specific expression generally conserved as well. For example, we and others found that the RNA-binding protein UNC-75/CELF is expressed in the nervous system of *C. elegans* where it controls the alternative splicing of hundreds of neuronal genes [82, 83, 114]. As is the case with many splicing factors, the UNC-75 family of RNA-binding proteins appears to have expanded significantly in mammals, which encode four homologues of UNC-75 [139]. Nevertheless, the human UNC-75 homologs (CELF3-6) are likewise expressed in the nervous system and affect alternative mRNA splicing [140, 141]. Likewise, EXC-7/ELAV RNA-binding protein is expressed in a subset of neurons in *C. elegans*, as are its mammalian splicing factor homologs [142, 143].

Similarly strikingly, recent large-scale profiling of *in vitro* binding sites for RNA-binding proteins throughout 24 eukaryotic species indicate that the *cis*-elements preferentially bound by splicing factors are highly conserved from worm to human [127]. For instance, the *C. elegans unc-75* consensus binding motif is a GU-rich sequence that is essentially identical to the consensus binding motifs of its homologues in *Drosophila* (Bru-3) and human (CELF4-6), and similar results were observed for other sets of homologous splicing factors [127].

Experiments in other systems have demonstrated that even though tissue-specific alternative splicing events tend to diverge over large enough timescales, the regulatory logic of individual tissue-specific RNA-binding proteins remains conserved. For example, the neuronal RNA-binding proteins NOVA1/2 in mammals and the *Drosophila* ortholog Pasilla not only recognize the same cis elements (YCAY motifs, in which Y is a pyrimidine), but operate with nearly identical "regulatory codes," for example: activating splicing of upstream exons and repressing slicing of downstream exons [122, 144]. Similarly, work comparing mouse and

human alternative splicing suggests that most species-specific alternative splicing is driven by changes in cis-elements, while the selectivity and activity of trans-acting splicing factors generally remain conserved [47]. These studies suggest that the mechanism and specificity of splicing factors is conserved over large evolutionary distances, and that rapid loss, gain and re-shuffling of cis-elements are responsible for the widespread divergence of alternative splicing among species. Future studies will determine whether these findings hold true in nematodes as well.

In sum, work performed in both mammals and invertebrates suggests that while individual tissue-specific alternative splicing events are poorly conserved across large evolutionary distances, the sequence-specificity of tissue-specific splicing factors is highly conserved. In the future, *C. elegans* should be a powerful system to look at the evolution of alternative splicing in greater depth, especially tissue-specific alternative splicing. Comparative studies can be performed on transgenic animals expressing fluorescent splicing reporters originating from different *Caenorhabditis* species, or reporters in which *cis*-elements from one species are substituted, and trans-acting splicing factors from other species can also be swapped between species. These comparative studies have been effective in analyzing the evolution of gene expression between mammals and vertebrates (for recent examples, see [47, 50, 145]) and between *Drosophilae* species [146]. Currently there are 10 reported *Caenorhabditis* species with genomes in various stages of assembly, and there are many more in progress (http://caenorhabditis.bio.ed.ac.uk/), which makes this genus an excellent resource for comparative evolutionary studies, including tissue-specific splicing, in genetically tractable organisms. Moreover, many new wild isolates of *C. elegans* from around the world, which have provided insights into selection pressures on alternative splicing [137], continue to be sequenced. This rich source of data should be valuable for understanding the evolution of tissue- and cell-specific alternative splicing.

12 Perspectives and Future Goals

12.1 Dynamic Regulation of Alternative Splicing in Response to Environmental Stimuli

To date, most studies of alternative splicing *in vivo* have focused on its role in development. However, several studies in mammalian cells have demonstrated that splicing can be dynamically regulated in response to stress or other stimuli [147–152]. These stimuli-induced splicing changes are of particular relevance in neurons, which frequently respond to differences in activity [147, 148, 153, 154]. A recent study in *C. elegans* has found that when animals recover from a starvation-induced larval arrest, significant changes in isoform usage occur as early as the first hour post-recovery [155]. These results suggest that nutrient availability significantly influence most layers of gene regulation, including splicing. More generally, *C. elegans*

will serve as an attractive model organism for *in vivo* studies of stimulus-induced regulation of RNA processing, because a number of behavioral paradigms have been established and studied. For instance, in addition to starvation-induced behaviors mentioned above, quality of diet can also significantly influence animal physiology and behavior [156–160]. Bacteria can simultaneously act as a food source, stimulus, and pathogen, and different bacterial strains elicit a wide variety of behavioral and physiological responses [161], and it will be interesting to see how gene regulatory networks and isoform usage change in response to different diets.

Additional environmental stimuli such as temperature, mechanical force, salt, and volatile or soluble chemicals have also been shown to elicit behavioral responses in animals [162, 163]. Research departing from investigations of animals growing solely under optimized laboratory conditions will reveal a wealth of mechanisms that have evolved to allow organisms to adapt to fluctuating environments, and how regulation of RNA processing can contribute to adaptation. Presumably, some of these adaptive mechanisms will have evolved in all metazoans, allowing molecular pathways in *C. elegans* to serve as tractable models for similar responses in vertebrates. Moreover, being able to easily study organism level adaptations to stimuli will lead to insights into how different tissue and cell types coordinate responses.

12.2 Splicing Regulation at the Level of Single Cells During Animal Development

C. elegans has proven to be an excellent model organism for addressing biological questions involving cell fate and differentiation. A number of studies investigating the role of transcription factors and chromatin modifiers in establishing cellular identity and function have taken advantage of the invariant cell lineage and wealth of molecular tools available in *C. elegans* to readily identify and label individual cells of interest (for recent examples see [164–166]). A number of studies have also shown that the loss of RNA-binding proteins can also significantly influence cellular identity and function, often impairing regulated splicing events [55, 56, 167–171].

Undoubtedly, concomitant with other layers of gene regulation, mRNA processing will also play an important role in shaping the transcriptomes of individual cells as they undergo non-symmetric and/or terminal divisions. Deciphering which of these RNA processing events have evolved to play a critical role in establishing and maintaining cellular identity, and the mechanisms governing these regulatory decisions will be a major goal moving forward. Studies making use of two color fluorescent reporters described above with genetic screens will continue to provide entry points into identifying key regulators of splicing events that define specific cell types. Complementary to this approach, the recent wealth of techniques for surveying the transcriptomes of populations of single cell types or even single cells [84, 172, 173] will provide a means to extend observations stemming from focused studies in a transcriptome-wide manner. Collectively, these experiments will be of particular importance in understanding how a nervous system is formed, which in *C. elegans* is

composed of over one hundred distinct classes of cells that express evolutionary conserved neurotransmitter systems.

Finally, computational approaches taking advantage of transcriptome-wide measurements of isoform abundances and protein-RNA interactomes show promise in identifying key sequence elements in the genome that dictate how tissue- and/or developmental stage-specific splicing is regulated [174–177]. Applying these computational approaches on datasets sampling single cell types or even single cells in *C. elegans* will allow for a greater appreciation and understanding of the 'splicing code' at cellular resolution. Ultimately, insights from tractable organisms like *C. elegans* will serve as paradigms for how processes like alternative splicing are integrated with other layers of gene regulation that contribute to development and differentiation in more complex metazoans.

Acknowledgments The authors would like to thank the Bauer Fellows Program, Harvard University and the NIH (Early Independence Award DP5OD09153) for continued support of our research.

References

1. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215–233
2. Hudson WH, Ortlund EA (2014) The structure, function and evolution of proteins that bind DNA and RNA. Nat Rev Mol Cell Biol 15:749–760
3. Quinodoz S, Guttman M (2014) Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends Cell Biol 24:651–663
4. Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ (2013) Dynamic integration of splicing within gene regulatory pathways. Cell 152:1252–1269
5. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463:457–463
6. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10:741–754
7. Schmucker D, Chen B (2009) Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. Genes Dev 23:147–156
8. Hattori D, Chen Y, Matthews BJ, Salwinski L, Sabatti C, Grueber WB, Zipursky SL (2009) Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature 461:644–648
9. Hattori D, Demir E, Kim HW, Viragh E, Zipursky SL, Dickson BJ (2007) Dscam diversity is essential for neuronal wiring and self-recognition. Nature 449:223–227
10. Matera AG, Wang Z (2014) A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15:108–121
11. Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718
12. Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ (2011) Ordered and dynamic assembly of single spliceosomes. Science 331:1289–1295
13. Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R (2006) Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res 34: 3955–3967
14. Chasin LA (2007) Searching for splicing motifs. Adv Exp Med Biol 623:85–106

15. Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14:802–813
16. Caceres JF, Stamm S, Helfman DM, Krainer AR (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science 265:1706–1709
17. Smith CW, Valcarcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci 25:381–388
18. Bradley T, Cook ME, Blanchette M (2015) SR proteins control a complex network of RNA-processing events. RNA 21:75–92
19. Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S et al (2012) Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1:167–178
20. Motta-Mena LB, Heyd F, Lynch KW (2010) Context-dependent regulatory mechanism of the splicing factor hnRNP L. Mol Cell 37:223–234
21. de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell 12:525–532
22. Allo M, Schor IE, Munoz MJ, de la Mata M, Agirre E, Valcarcel J, Eyras E, Kornblihtt AR (2010) Chromatin and alternative splicing. Cold Spring Harb Symp Quant Biol 75:103–111
23. Haque N, Oberdoerffer S (2014) Chromatin and splicing. Methods Mol Biol 1126:97–113
24. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. Cell 144:16–26
25. Schwartz S, Ast G (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. EMBO J 29:1629–1636
26. McManus CJ, Graveley BR (2011) RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev 21:373–379
27. Shepard PJ, Hertel KJ (2008) Conserved RNA secondary structures promote alternative splicing. RNA 14:1463–1469
28. Luco RF, Misteli T (2011) More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. Curr Opin Genet Dev 21:366–372
29. Grosso AR, Gomes AQ, Barbosa-Morais NL, Caldeira S, Thorne NP, Grech G, von Lindern M, Carmo-Fonseca M (2008) Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res 36:4823–4832
30. Hanamura A, Caceres JF, Mayeda A, Franza BR Jr, Krainer AR (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA 4:430–444
31. Zhou Z, Fu XD (2013) Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma 122:191–207
32. Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, Okano HJ, Yang YY, Darnell RB (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron 25:359–371
33. Kalsotra A, Xiao X, Ward AJ, Castle JC, Johnson JM, Burge CB, Cooper TA (2008) A post-natal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proc Natl Acad Sci U S A 105:20333–20338
34. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP (2009) ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol Cell 33:591–601
35. Calarco JA, Zhen M, Blencowe BJ (2011) Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. RNA 17:775–791
36. Darnell RB (2013) RNA protein interaction in neurons. Annu Rev Neurosci 36:243–270
37. Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M Jr, Black DL (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev 21:1636–1652
38. Spellman R, Llorian M, Smith CW (2007) Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Mol Cell 27:420–434

39. Makeyev EV, Zhang J, Carrasco MA, Maniatis T (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell 27:435–448
40. Li Q, Zheng S, Han A, Lin CH, Stoilov P, Fu XD, Black DL (2014) The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. eLife 3, e01201
41. Licatalosi DD, Yano M, Fak JJ, Mele A, Grabinski SE, Zhang C, Darnell RB (2012) Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. Genes Dev 26:1626–1642
42. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL (2000) Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol 20:7463–7479
43. Zheng S, Gray EE, Chawla G, Porse BT, O'Dell TJ, Black DL (2012) PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. Nat Neurosci 15(381–388):S381
44. Calarco JA, Superina S, O'Hanlon D, Gabut M, Raj B, Pan Q, Skalska U, Clarke L, Gelinas D, van der Kooy D et al (2009) Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. Cell 138:898–910
45. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallieres M, Tapial J, Raj B, O'Hanlon D et al (2014) A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 159:1511–1523
46. Raj B, Irimia M, Braunschweig U, Sterne-Weiler T, O'Hanlon D, Lin ZY, Chen GI, Easton LE, Ule J, Gingras AC et al (2014) A global regulatory mechanism for activating an exon network required for neurogenesis. Mol Cell 56:90–103
47. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R et al (2012) The evolutionary landscape of alternative splicing in vertebrate species. Science 338:1587–1593
48. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM et al (2014) Diversity and dynamics of the Drosophila transcriptome. Nature 512:393–399
49. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K et al (2010) Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330:1775–1787
50. Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science 338:1593–1599
51. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40:1413–1415
52. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C (2007) Rapid, transcript-specific changes in splicing in response to environmental stress. Mol Cell 27:928–937
53. Ramani AK, Calarco JA, Pan Q, Mavandadi S, Wang Y, Nelson AC, Lee LJ, Morris Q, Blencowe BJ, Zhen M et al (2011) Genome-wide analysis of alternative splicing in Caenorhabditis elegans. Genome Res 21:342–348
54. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476
55. Han H, Irimia M, Ross PJ, Sung HK, Alipanahi B, David L, Golipour A, Gabut M, Michael IP, Nachman EN et al (2013) MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. Nature 498:241–245
56. Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O'Hanlon D, Sung HK, Alvarez M, Talukder S, Pan Q, Mazzoni EO et al (2011) An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. Cell 147:132–146
57. Kalsotra A, Wang K, Li PF, Cooper TA (2010) MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. Genes Dev 24:653–658
58. Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, Guo W, Xing Y, Carstens RP (2010) An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. EMBO J 29:3286–3300

59. Raj B, O'Hanlon D, Vessey JP, Pan Q, Ray D, Buckley NJ, Miller FD, Blencowe BJ (2011) Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. Mol Cell 43:843–850
60. Elkon R, Ugalde AP, Agami R (2013) Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet 14:496–506
61. Smith L (2008) Post-transcriptional regulation of gene expression by alternative 5′-untranslated regions in carcinogenesis. Biochem Soc Trans 36:708–711
62. Popp MW, Maquat LE (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. Annu Rev Genet 47:139–165
63. Lareau LF, Brooks AN, Soergel DA, Meng Q, Brenner SE (2007) The coupling of alternative splicing and nonsense-mediated mRNA decay. Adv Exp Med Biol 623:190–211
64. Barberan-Soler S, Lambert NJ, Zahler AM (2009) Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. RNA 15:1652–1660
65. Barberan-Soler S, Zahler AM (2008) Alternative splicing regulation during C elegans development: splicing factors as regulated targets. PLoS Genet 4, e1000001
66. Jangi M, Boutz PL, Paul P, Sharp PA (2014) Rbfox2 controls autoregulation in RNA-binding protein networks. Genes Dev 28:637–651
67. Jangi M, Sharp PA (2014) Building robust transcriptomes with master splicing factors. Cell 159:487–498
68. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446:926–929
69. Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M Jr (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 21:708–718
70. Saltzman AL, Pan Q, Blencowe BJ (2011) Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev 25:373–384
71. Blumenthal T, Steward K (1997) RNA processing and gene structure. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR (eds) C elegans II. Cold Spring Harbor, New York
72. Zahler AM (2012) Pre-mRNA splicing and its regulation in Caenorhabditis elegans. WormBook: the online review of C elegans biology. 1–21
73. Blumenthal T (2012) Trans-splicing and operons in C. elegans. WormBook: the online review of C elegans biology, pp 1–11
74. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO (2014) Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Mol Biol Evol 31:1402–1413
75. Alfonso A, Grundahl K, McManus JR, Asbury JM, Rand JB (1994) Alternative splicing leads to two cholinergic proteins in Caenorhabditis elegans. J Mol Biol 241:627–630
76. Gross RE, Bagchi S, Lu X, Rubin CS (1990) Cloning, characterization, and expression of the gene for the catalytic subunit of cAMP-dependent protein kinase in Caenorhabditis elegans. Identification of highly conserved and unique isoforms generated by alternative splicing. J Biol Chem 265:6896–6907
77. Lee MH, Jang YJ, Koo HS (1998) Alternative splicing in the Caenorhabditis elegans DNA topoisomerase I gene. Biochim Biophys Acta 1396:207–214
78. Lundquist EA, Herman RK, Rogalski TM, Mullen GP, Moerman DG, Shaw JE (1996) The mec-8 gene of C. elegans encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts. Development 122:1601–1610
79. Sibley MH, Johnson JJ, Mello CC, Kramer JM (1993) Genetic identification, sequence, and alternative splicing of the Caenorhabditis elegans alpha 2(IV) collagen gene. J Cell Biol 123:255–264
80. Kent WJ, Zahler AM (2000) The intronerator: exploring introns and alternative splicing in Caenorhabditis elegans. Nucleic Acids Res 28:91–93
81. Barberan-Soler S, Medina P, Estella J, Williams J, Zahler AM (2011) Co-regulation of alternative splicing by diverse splicing factors in Caenorhabditis elegans. Nucleic Acids Res 39:666–674

82. Kuroyanagi H, Watanabe Y, Suzuki Y, Hagiwara M (2013) Position-dependent and neuron-specific splicing regulation by the CELF family RNA-binding protein UNC-75 in Caenorhabditis elegans. Nucleic Acids Res 41:4015–4025
83. Norris AD, Gao S, Norris ML, Ray D, Ramani AK, Fraser AG, Morris Q, Hughes TR, Zhen M, Calarco JA (2014) A pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. Mol Cell 54:946–959
84. Spencer WC, McWhirter R, Miller T, Strasbourger P, Thompson O, Hillier LW, Waterston RH, Miller DM 3rd (2014) Isolation of specific neurons from C elegans larvae for gene expression profiling. PLoS One 9, e112102
85. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, McWhirter RD, Petersen S, Sreedharan VT, Widmer C, Jo J et al (2011) A spatial and temporal map of C. elegans gene expression. Genome Res 21:325–341
86. Zhang Y, Ma C, Delohery T, Nasipak B, Foat BC, Bounoutas A, Bussemaker HJ, Kim SK, Chalfie M (2002) Identification of genes expressed in C. elegans touch receptor neurons. Nature 418:331–335
87. Roy PJ, Stuart JM, Lund J, Kim SK (2002) Chromosomal clustering of muscle-expressed genes in Caenorhabditis elegans. Nature 418:975–979
88. Von Stetina SE, Watson JD, Fox RM, Olszewski KL, Spencer WC, Roy PJ, Miller DM 3rd (2007) Cell-specific microarray profiling experiments reveal a comprehensive picture of gene expression in the C elegans nervous system. Genome Biol 8:R135
89. Kuroyanagi H, Ohno G, Sakane H, Maruoka H, Hagiwara M (2010) Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic Caenorhabditis elegans. Nat Protoc 5:1495–1517
90. Orengo JP, Bundman D, Cooper TA (2006) A bichromatic fluorescent reporter for cell-based screens of alternative splicing. Nucleic Acids Res 34, e148
91. Kuroyanagi H, Kobayashi T, Mitani S, Hagiwara M (2006) Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. Nat Methods 3:909–915
92. Goodman SJ, Branda CS, Robinson MK, Burdine RD, Stern MJ (2003) Alternative splicing affecting a novel domain in the C. elegans EGL-15 FGF receptor confers functional specificity. Development 130:3757–3766
93. Ohno G, Ono K, Togo M, Watanabe Y, Ono S, Hagiwara M, Kuroyanagi H (2012) Muscle-specific splicing factors ASD-2 and SUP-12 cooperatively switch alternative pre-mRNA processing patterns of the ADF/cofilin gene in Caenorhabditis elegans. PLoS Genet 8, e1002991
94. Pujol N, Bonnerot C, Ewbank JJ, Kohara Y, Thierry-Mieg D (2001) The Caenorhabditis elegans unc-32 gene encodes alternative forms of a vacuolar ATPase a subunit. J Biol Chem 276:11913–11921
95. Kuroyanagi H, Watanabe Y, Hagiwara M (2013) CELF family RNA-binding protein UNC-75 regulates two sets of mutually exclusive exons of the unc-32 gene in neuron-specific manners in Caenorhabditis elegans. PLoS Genet 9, e1003337
96. Ono K, Parast M, Alberico C, Benian GM, Ono S (2003) Specific requirement for two ADF/cofilin isoforms in distinct actin-dependent processes in Caenorhabditis elegans. J Cell Sci 116:2073–2085
97. Anyanful A, Ono K, Johnsen RC, Ly H, Jensen V, Baillie DL, Ono S (2004) The RNA-binding protein SUP-12 controls muscle-specific splicing of the ADF/cofilin pre-mRNA in C. elegans. J Cell Biol 167:639–647
98. Lo TW, Branda CS, Huang P, Sasson IE, Goodman SJ, Stern MJ (2008) Different isoforms of the C. elegans FGF receptor are required for attraction and repulsion of the migrating sex myoblasts. Dev Biol 318:268–275
99. Bulow HE, Boulin T, Hobert O (2004) Differential functions of the C. elegans FGF receptor in axon outgrowth and maintenance of axon position. Neuron 42:367–374
100. Iijima T, Iijima Y, Witte H, Scheiffele P (2014) Neuronal cell type-specific alternative splicing is regulated by the KH domain protein SLM1. J Cell Biol 204:331–342

101. Schreiner D, Nguyen TM, Russo G, Heber S, Patrignani A, Ahrne E, Scheiffele P (2014) Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. Neuron 84:386–398
102. Miura SK, Martins A, Zhang KX, Graveley BR, Zipursky SL (2013) Probabilistic splicing of Dscam1 establishes identity at the level of single neurons. Cell 155:1166–1177
103. Pinan-Lucarre B, Tu H, Pierron M, Cruceyra PI, Zhan H, Stigloher C, Richmond JE, Bessereau JL (2014) C. elegans Punctin specifies cholinergic versus GABAergic identity of postsynaptic domains. Nature 511:466–470
104. Ohno H, Kato S, Naito Y, Kunitomo H, Tomioka M, Iino Y (2014) Role of synaptic phosphatidylinositol 3-kinase in a behavioral learning response in C. elegans. Science 345:313–317
105. Kuroyanagi H, Ohno G, Mitani S, Hagiwara M (2007) The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. Mol Cell Biol 27:8612–8621
106. Amrane S, Rebora K, Zniber I, Dupuy D, Mackereth CD (2014) Backbone-independent nucleic acid binding by splicing factor SUP-12 reveals key aspects of molecular recognition. Nat Commun 5:4595
107. Kuwasako K, Takahashi M, Unzai S, Tsuda K, Yoshikawa S, He F, Kobayashi N, Guntert P, Shirouzu M, Ito T et al (2014) RBFOX and SUP-12 sandwich a G base to cooperatively regulate tissue-specific splicing. Nat Struct Mol Biol 21:778–786
108. Ohno G, Hagiwara M, Kuroyanagi H (2008) STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. Genes Dev 22:360–374
109. Galarneau A, Richard S (2005) Target RNA motif and target mRNAs of the Quaking STAR protein. Nat Struct Mol Biol 12:691–698
110. Pandya-Jones A, Black DL (2009) Co-transcriptional splicing of constitutive and alternative exons. RNA 15:1896–1908
111. Byrd DT, Kawasaki M, Walcoff M, Hisamoto N, Matsumoto K, Jin Y (2001) UNC-16, a JNK-signaling scaffold protein, regulates vesicle transport in C. elegans. Neuron 32:787–800
112. Edwards SL, Yu SC, Hoover CM, Phillips BC, Richmond JE, Miller KG (2013) An organelle gatekeeper function for Caenorhabditis elegans UNC-16 (JIP3) at the axon initial segment. Genetics 194:143–161
113. Fujita M, Kawano T, Ohta A, Sakamoto H (1999) Neuronal expression of a Caenorhabditis elegans elav-like gene and the effect of its ectopic expression. Biochem Biophys Res Commun 260:646–652
114. Loria PM, Duke A, Rand JB, Hobert O (2003) Two neuronal, nuclear-localized RNA-binding proteins involved in synaptic transmission. Curr Biol 13:1317–1323
115. Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela SA et al (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol 16:717–724
116. Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau JC et al (2012) Argonaute proteins couple chromatin silencing to alternative splicing. Nat Struct Mol Biol 19:998–1004
117. Barberan-Soler S, Fontrodona L, Ribo A, Lamm AT, Iannone C, Ceron J, Lehner B, Valcarcel J (2014) Co-option of the piRNA pathway for germline-specific alternative splicing of C. elegans TOR. Cell Rep 8:1609–1616
118. Kabat JL, Barberan-Soler S, McKenna P, Clawson H, Farrer T, Zahler AM (2006) Intronic alternative splicing regulators identified by comparative genomics in nematodes. PLoS Comput Biol 2, e86
119. Kabat JL, Barberan-Soler S, Zahler AM (2009) HRP-2, the Caenorhabditis elegans homolog of mammalian heterogeneous nuclear ribonucleoproteins Q and R, is an alternative splicing factor that binds to UCUAUC splicing regulatory elements. J Biol Chem 284:28490–28497
120. Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

121. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
122. Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB (2006) An RNA map predicting Nova-dependent splicing regulation. Nature 444:580–586
123. Jungkamp AC, Stoeckius M, Mecenas D, Grun 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
124. Pagano JM, Farley BM, Essien KI, Ryder SP (2009) RNA recognition by the embryonic cell fate determinant and germline totipotency factor MEX-3. Proc Natl Acad Sci U S A 106:20252–20257
125. 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
126. 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
127. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, Gueroussov S, Albu M, Zheng H, Yang A et al (2013) A compendium of RNA-binding motifs for decoding gene regulation. Nature 499:172–177
128. 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
129. Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpaa MJ et al (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res 20:861–873
130. Zykovich A, Korf I, Segal DJ (2009) Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. Nucleic Acids Res 37:e151
131. Malko DB, Makeev VJ, Mironov AA, Gelfand MS (2006) Evolution of exon-intron structure and alternative splicing in fruit flies and malarial mosquito genomes. Genome Res 16:505–509
132. Modrek B, Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34:177–180
133. Irimia M, Rukov JL, Penny D, Garcia-Fernandez J, Vinther J, Roy SW (2008) Widespread evolutionary conservation of alternatively spliced exons in Caenorhabditis. Mol Biol Evol 25:375–382
134. Fagnani M, Barash Y, Ip JY, Misquitta C, Pan Q, Saltzman AL, Shai O, Lee L, Rozenhek A, Mohammad N et al (2007) Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biol 8:R108
135. Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11:345–355
136. Sugnet CW, Srinivasan K, Clark TA, O'Brien G, Cline MS, Wang H, Williams A, Kulp D, Blume JE, Haussler D et al (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput Biol 2, e4
137. Barberan-Soler S, Zahler AM (2008) Alternative splicing and the steady-state ratios of mRNA isoforms generated by it are under strong stabilizing selection in Caenorhabditis elegans. Mol Biol Evol 25:2431–2437
138. Rukov JL, Irimia M, Mork S, Lund VK, Vinther J, Arctander P (2007) High qualitative and quantitative conservation of alternative splicing in Caenorhabditis elegans and Caenorhabditis briggsae. Mol Biol Evol 24:909–917
139. Barbosa-Morais NL, Carmo-Fonseca M, Aparicio S (2006) Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. Genome Res 16:66–77
140. Dasgupta T, Ladd AN (2012) The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. Wiley Interdiscip Rev RNA 3:104–121

141. Ladd AN, Charlet N, Cooper TA (2001) The CELF family of RNA-binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol 21:1285–1296
142. Hinman MN, Lou H (2008) Diverse molecular functions of Hu proteins. Cell Mol Life Sci 65:3168–3181
143. Ince-Dunn G, Okano HJ, Jensen KB, Park WY, Zhong R, Ule J, Mele A, Fak JJ, Yang C, Zhang C et al (2012) Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. Neuron 75:1067–1080
144. Brooks AN, Yang L, Duff MO, Hansen KD, Park JW, Dudoit S, Brenner SE, Graveley BR (2011) Conservation of an RNA regulatory map between Drosophila and mammals. Genome Res 21:193–202
145. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, Dobin A, Zaleski C, Beer MA, Chapman WC et al (2014) Comparison of the transcriptional landscapes between human and mouse tissues. Proc Natl Acad Sci U S A 111:17224–17229
146. Naval-Sanchez M, Potier D, Haagen L, Sanchez M, Munck S, Van de Sande B, Casares F, Christiaens V, Aerts S (2013) Comparative motif discovery combined with comparative transcriptomics yields accurate targetome and enhancer predictions. Genome Res 23:74–88
147. An P, Grabowski PJ (2007) Exon silencing by UAGG motifs in response to neuronal excitation. PLoS Biol 5, e36
148. Eom T, Zhang C, Wang H, Lay K, Fak J, Noebels JL, Darnell RB (2013) NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. eLife 2, e00178
149. Ip JY, Schmidt D, Pan Q, Ramani AK, Fraser AG, Odom DT, Blencowe BJ (2011) Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. Genome Res 21:390–401
150. Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, Glover-Cutter K, Ben-Dov C, Blaustein M, Lozano JJ et al (2009) DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell 137:708–720
151. Shalgi R, Hurt JA, Lindquist S, Burge CB (2014) Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. Cell Rep 7:1362–1370
152. Xie J, Black DL (2001) A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. Nature 410:936–939
153. Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60:1022–1038
154. Lee JA, Xing Y, Nguyen D, Xie J, Lee CJ, Black DL (2007) Depolarization and CaM kinase IV modulate NMDA receptor splicing through two essential RNA elements. PLoS Biol 5, e40
155. Maxwell CS, Antoshechkin I, Kurhanewicz N, Belsky JA, Baugh LR (2012) Nutritional control of mRNA isoform expression during developmental arrest and recovery in C. elegans. Genome Res 22:1920–1929
156. Cabreiro F, Au C, Leung KY, Vergara-Irigaray N, Cocheme HM, Noori T, Weinkove D, Schuster E, Greene ND, Gems D (2013) Metformin retards aging in C. elegans by altering microbial folate and methionine metabolism. Cell 153:228–239
157. Kim DH, Feinbaum R, Alloing G, Emerson FE, Garsin DA, Inoue H, Tanaka-Hino M, Hisamoto N, Matsumoto K, Tan MW et al (2002) A conserved p38 MAP kinase pathway in Caenorhabditis elegans innate immunity. Science 297:623–626
158. MacNeil LT, Watson E, Arda HE, Zhu LJ, Walhout AJ (2013) Diet-induced developmental acceleration independent of TOR and insulin in C. elegans. Cell 153:240–252
159. Shtonda BB, Avery L (2006) Dietary choice behavior in Caenorhabditis elegans. J Exp Biol 209:89–102
160. Zhang Y, Lu H, Bargmann CI (2005) Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. Nature 438:179–184

161. Macneil LT, Walhout AJ (2013) Food, pathogen, signal: the multifaceted nature of a bacterial diet. Worm 2, e26454
162. Bargmann CI (2006) Chemosensation in C. elegans. WormBook: the online review of C elegans biology. 1–29
163. Goodman MB (2006) Mechanosensation. WormBook: the online review of C elegans biology. 1–14
164. Araya CL, Kawli T, Kundaje A, Jiang L, Wu B, Vafeados D, Terrell R, Weissdepp P, Gevirtzman L, Mace D et al (2014) Regulatory analysis of the C. elegans genome with spatio-temporal resolution. Nature 512:400–405
165. Serrano-Saiz E, Poole RJ, Felton T, Zhang F, De La Cruz ED, Hobert O (2013) Modular control of glutamatergic neuronal identity in C. elegans by distinct homeodomain proteins. Cell 155:659–673
166. Zuryn S, Ahier A, Portoso M, White ER, Morin MC, Margueron R, Jarriault S (2014) Transdifferentiation. Sequential histone-modifying activities determine the robustness of transdifferentiation. Science 345:826–829
167. Gehman LT, Meera P, Stoilov P, Shiue L, O'Brien JE, Meisler MH, Ares M Jr, Otis TS, Black DL (2012) The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. Genes Dev 26:445–460
168. Huang CS, Shi SH, Ule J, Ruggiu M, Barker LA, Darnell RB, Jan YN, Jan LY (2005) Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell 123:105–118
169. Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW, Swanson MS (2003) A muscleblind knockout model for myotonic dystrophy. Science 302:1978–1980
170. Venables JP, Lapasset L, Gadea G, Fort P, Klinck R, Irimia M, Vignal E, Thibault P, Prinos P, Chabot B et al (2013) MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. Nat Commun 4:2480
171. Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol 16:130–137
172. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2:666–673
173. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098
174. Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. Nature 465:53–59
175. Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, Xue C, Herre M, Silver PA, Zhang MQ et al (2014) HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep 6:1139–1152
176. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RK, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR et al (2014) The human splicing code reveals new insights into the genetic determinants of disease. Science 347(6218):1254806
177. Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science 329:439–443

Chapter 11
RNA Granules and Diseases: A Case Study of Stress Granules in ALS and FTLD

Alexander C. Fan and Anthony K.L. Leung

Abstract RNA granules are microscopically visible cellular structures that aggregate by protein–protein and protein–RNA interactions. Using stress granules as an example, we discuss the principles of RNA granule formation, which rely on the multivalency of RNA and multi-domain proteins as well as low-affinity interactions between proteins with prion-like/low-complexity domains (e.g. FUS and TDP-43). We then explore how dysregulation of RNA granule formation is linked to neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD), and discuss possible strategies for therapeutic intervention.

Keywords Stress granules • Low-complexity region • Prion-like domain • RNA granules • Phase separation

1 A Cellular World of RNA Granules

Cellular RNAs rarely act alone. Rather, their functions are mediated through RNA-binding proteins (RBPs) and other functional partners, resulting in the formation of ribonucleoprotein (RNP) complexes. MicroRNAs, for example, associate with their cognate binding partner Argonaute, which, in turn, serves as a platform to recruit other proteins to form the silencing complexes responsible for translation repression and/or decay of mRNA targets [1, 2]. Besides microRNA-induced silencing complexes, various types of other RNPs are formed to regulate gene expression over the life cycle of an mRNA. For example, the birth of a eukaryotic mRNA is mediated by a cascade of proteins that cap the 5′ end, add a poly(A) tail at the 3′ end, and carry out splicing to remove introns. The process generally terminates with the deposition of exon junction complexes following splicing, which signal for the mature transcript to be exported to

A.C. Fan • A.K.L. Leung (✉)
Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
e-mail: anthony.leung@jhu.edu

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_11

the cytoplasm where it can be bound by ribosomes for translation or exosomes for degradation [3]. Whether the mature mRNA is destined for translation or degradation is, in turn, determined by RBPs bound to the transcript, often located in the 3′ untranslated regions (UTRs) [4]. Thus, from transcription in the nucleus to degradation in the cytoplasm, proteins of distinct function often associate as macrocomplexes of RNPs to fulfill essential regulatory functions. Strikingly, RNPs frequently undergo further aggregation to form microscopically visible structures commonly termed RNA granules [5–7]. These structures can be nuclear or cytoplasmic, and their mechanisms of action are only just beginning to be understood. Classes of these structures include nucleoli, Cajal bodies, nuclear speckles, and paraspeckles in the nucleus, as well as P-bodies and stress granules in the cytoplasm [8].

A striking feature of RNA granules as a structural class is that they maintain defined protein and RNA compositions in the absence of an enveloping membrane. In this review, we will focus on the potential mechanism(s) for the formation and regulation of a specific class of cytoplasmic RNA granule known as the stress granule (SG). Unlike other RNA granules, these cytoplasmic structures are not constitutively present; instead, their formation is induced by cellular stress, such as heat shock or oxidative stress, and they disassemble once the perturbation subsides. Notably, morphologically similar cytoplasmic inclusions are observed in neurons of patients with amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD) and other neurodegenerative disease, often exhibiting compositional overlap with endogenous SGs. These morphological phenotypes suggest that the formation or disassembly process of these structures may be important for neurological pathogenesis. We will conclude this chapter by focusing on the pathogenic roles of two RBPs—FUS and TDP-43—and how their mutations link dysregulation of stress granules to ALS/FTLD.

2 Stress Granules

Stress granules (SGs) were first discovered in mammalian cells [9–12], and similar structures were later identified in *Drosphila* and yeast cells [13–16]. Electron microscopy revealed that mammalian SGs are spheroid or ellipsoid structures that are usually 1–2 μm, but can range from 0.4 to 5 μm [8, 17]. SGs can be induced upon a variety of stresses that inhibit translation initiation, including heat shock, oxidative stress, hypoxia, inhibition of mitochondrial function, glucose starvation, proteotoxic stress, and infection by certain viruses [18–21]. Stalled protein synthesis under these conditions is evident by the concomitant loss of conventional "polysomal rosettes" [17, 22]. Although SGs are intimately tied to polysome disassembly, translation inhibition is not in itself sufficient to induce their aggregation since pharmacological inhibition of 60S subunit recruitment can halt translation without SG formation [23]. Rather, a defining feature for mammalian SGs has been the presence of stalled 48S pre-initiation complex subunits [6]. In the following sub-section, we proceed with an introduction to the molecular composition and proposed functions of SGs.

2.1 Molecular Composition

Inhibition of translation initiation is the trigger for SG formation, and subcomponents of the 48S complex—including eukaryotic initiation factors eIF3, eIF4F, eIF4B, the small ribosomal subunits, poly(A) binding protein PABP1, and stalled mRNA transcripts—likely initiate SG assembly through a series of concerted protein–protein and protein–RNA interactions. These macromolecules comprise the core class of SG components [18] (Fig. 11.1).

Additional protein components are incorporated into SGs after the primary initating event, including a large number of other RBPs, many of which are known regulators of mRNA translation and stability [6]. Intriguingly, proteins within this class, including TIA-1, G3BP1, CPEB, FXR1, FMRP, and MLN51 [12, 24–27], can induce spontaneous formation of SGs when overexpressed even in the absence of stress. Many proteins in this class, such as TIA-1 and G3BP1, are often used as protein markers to identify SGs using fluorescence microscopy [19]. A final class of SG components consists of tertiary "piggy-back" components, which might not function in RNA metabolism *per se*, but are recruited into SGs by protein–protein interactions [18]. Recent exploration of these proteins, DYRK3 and RACK1 being notable examples, has suggested expanded functions for SGs in a variety of cell-signaling pathways including apoptosis and the mTOR pathway [28] (see Sect. 11.2.2).

It is important to note that the composition of SGs is variable and depends on the type of stress used to induce them. For example, heat shock protein Hsp27 is found in

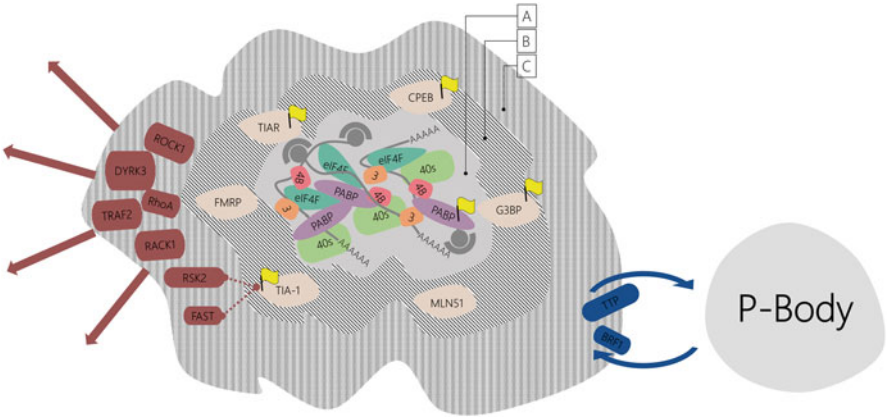


Fig. 11.1 Classes of SG components. (**a**) The defining first class of SG components includes subcomponents of the 48S pre-initiation complex, along with stalled mRNA transcripts. (**b**) A second class of proteins that either bind mRNA directly or are partners of RBPs. Many of these proteins contain unstructured domains and can nucleate SG assembly. Some are commonly used as "markers" to detect SGs by immunofluorescence (*yellow flags*). (**c**) Tertiary "piggy back" proteins, many of which function in cell signaling pathways (*red arrows*)

SGs in heat-shocked mammalian cells, but not in cells subjected to arsenite treatment [12], while poliovirus-induced SGs contain a unique marker, Sam68 [29]. Similarly, in yeast, SGs induced by glucose deprivation contain the heterogeneous nuclear ribonucleoprotein (hnRNP) Hrp1, whereas those that form under inhibition of mitochondrial respiration do not [30]. Finally, while yeast SGs formed during robust heat shock at 46 °C for 10 min contain the canonical components eIF3 and the 40S ribosomal units [15], those formed under glucose-deprived conditions, surprisingly, do not [13, 16]. These data suggest that, rather than a series of defined binding events proceeding orderly from a nucleating mRNP, SGs more likely exist along a continuum of compositional states dependent on the pathways activated by specific stress conditions [7]. For example, TTP is removed from SGs when phosphorylated by MK2 [31]. Thus, SG composition depends on how these stressors cause the redistribution of various mRNPs between SGs and translation-competent polysomes, as well as between SGs and P-bodies (PBs)—another class of cytoplasmic RNA granule enriched for RNA decay factors. As will be emphasized in subsequent sections, SGs house mechanisms to selectively exclude certain RNA transcripts so as to prioritize translation of protein chaperones [32, 33]. Besides these pathways, however, little is known about the identity of other RNA components, whether their recruitment is regulated in a stress-dependent manner, and how this might influence protein composition.

2.2 Proposed Functions

Though a global proteomics approach has not yet been used to assess the full extent of proteins recruited into SGs (see "Notes" at the end of the Chapter), new components continue to be identified by co-localization studies using defined protein markers and fluorescence microscopy. Since these new components are not necessarily related to RNA metabolism, the proposed functions for SGs are rapidly expanding. Given the presence of translationally stalled mRNA complexes, SGs were originally hypothesized to function as storage sites for non-translating RNAs. However, fluorescence recovery after photobleaching (FRAP) analyses showed that both mRNA transcripts and nearly all protein components shuttle in and out of SGs with half-lives on the order of seconds to minutes [25, 34–36]. Thus, it is difficult to envisage that SGs act as static mRNA repositories. Instead, an emerging, working model for SG function posits that they are sites of dynamic mRNP remodeling and sorting, a process which has been coined mRNA triage [37]. Under this framework, SGs function as transient structures induced by perturbations to translation initiation, during which they exchange components with polysomes and PBs to prioritize translation or degradation of some mRNA transcripts over others, thereby altering the proteome until stress subsides.

In support of this, strong evidence exists for a direct flow of mRNPs between polysomes and SGs. First, puromycin treatment, which releases ribosomes from translating mRNAs by inducing premature termination of translation, enhances SG formation [38]. Emetine, which functions to the opposite effect by preventing mRNA-ribosome dissociation, both inhibits the formation of SGs in the presence of

stress (e.g. sodium arsenite treatment) and even "dissolves" pre-formed SGs [38]. Combined with the quick turnover rates as assessed by FRAP, these data suggest that SGs are are constantly "fed" by translationally stalled mRNAs originating from polysomes. Moreover, a specific pathway of mRNP trafficking between SGs and polysomes was recently characterized in yeast, facilitated by the ATP-dependent RNA helicase Ded1 [39]. Ded1 binds eIF4G1 and interrupts translation, forming a translationally inactive pre-initiation complex with eIF4F and stalled mRNA which then accumulates in SGs [39]. Upon ATP hydrolysis, Ded1 is thought to release the pre-initiation mRNP, regulating its return to translation [39]. The mammalian orthologue DDX3 has been found to interact with eIF4E, a binding partner of eIF4G in the mRNA cap structure [40], though it remains unclear whether it carries specific roles in polysome/SG trafficking.

Another corner of this triage model is grounded in interactions between SGs and PBs. In a subset of cellular stress models, SGs and PBs have been observed to physically interact. Under sodium arsenite treatment, for example, PBs can make contact with SGs while remaining morphologically distinct [17], or may otherwise exist in close proximity with them [25, 27]. In a series of experiments utilizing time-lapse microscopy in conjunction with transient expression of tagged SG or PB protein markers, interactions between the two bodies were observed to vary — PBs appeared stably bound to SGs in some cases, whereas in others they were only intermittently attached [25]. These interactions, which might only be observed in a fraction of imaged cells, can be significantly enhanced by overexpression of certain shared proteins. Most notably, transient expression of the mRNA destabilizing factors TTP and BRF1 were shown to induce quantitative SG–PB docking, extending the tethered interaction to 1 h [25]. Similarly, overexpression of CPEB1, a protein observed to localize to both types of RNA granules, resulted in PB clustering around spontaneously induced SGs [27].

Beyond the mRNA triage model, an additional key functional property of SGs is the acute and localized enrichment of proteins responsible for RNA metabolism and various signaling pathways. Indeed, dozens of different proteins have been found to localize to SGs [18], and the local concentration of these components may either serve to increase the rate of biochemical reactions within SGs or reduce the cytosolic concentration of other proteins by sequestration, resulting in spillover effects that propagate through a variety of cell signaling pathways [28]. For example, a number of emerging lines of evidence strongly suggest that one major function of SGs is to regulate apoptosis [28, 41–44]. First, SGs sequester pro-apoptotic factors such as RACK1 and TRAF2 from the cytosol to suppress the activation of cellular death pathways via MAPK or NF-kB signaling [41, 45]. Similarly, the sequestration of small GTPase RhoA and its downstream kinase ROCK1 by SGs prevents the activation of JNK which would otherwise trigger apoptosis [46]. Second, SGs recruit anti-apoptotic factors, such as RSK2 and FAST, which bind and inactivate apoptosis promoting factor TIA-1 [42, 47]. Third, SG formation stabilizes mRNAs encoding anti-apoptotic protein $p21^{WAF1/CIP1}$ [43]. Consistent with these findings, impairment of SG assembly often leads to a higher rate of cell death following stress exposure [24, 42, 48]. In addition, SG formation appears to regulate the mTOR pathway, which monitors nutrient and energy availability to promote either cell growth when conditions are favorable or catabolic processes during stress [44, 49–51]. In this case, the

kinase DYRK3, which itself is recruited into SGs, phosphorylates the mTORC1 inhibitor PRAS40, allowing mTORC1 to exit SGs in an activated state to fulfill its signaling functions in cell growth and metabolism [51].

In summary, SGs are implicated as regulatory hubs for mRNA sorting, stress signaling, and apoptosis. Besides various physiological stresses, SGs can also be induced by radiotherapy and chemotherapeutics, as is the case with the proteasome inhibitor Velcade in solid tumors [20, 52]. The persistent presence of cytoplasmic inclusions containing SG components in hypoxic tumor cores and ALS/FTLD patient neurons has been linked to chemotherapy resistance in cancers and neurodegeneration, respectively [41, 53]. On the other hand, SGs seem to also serve in antiviral functions given that various viruses inhibit SG formation immediately upon infection [21, 54]. Thus, it is critical to identify the molecular mechanisms that hold SGs together in physiological and pathophysiological contexts. In the next section, we will focus on a molecular model of how non-membranous RNA granules can be formed in a way that allows for select proteins to be retained while others move in and out readily.

3 RNA Granule Formation via Multivalency and Intrinsically Disordered Regions

An evolving framework of RNA granule formation, which works to explain assembly, regulation, and selectivity in recruitment, is grounded in the biophysical mechanisms facilitated by two properties: (1) multivalency mediated by proteins or RNA, and (2) low affinity protein–protein interactions involving intrinsically disordered regions of low complexity amino acid composition (i.e., a string of repetitive amino acids).

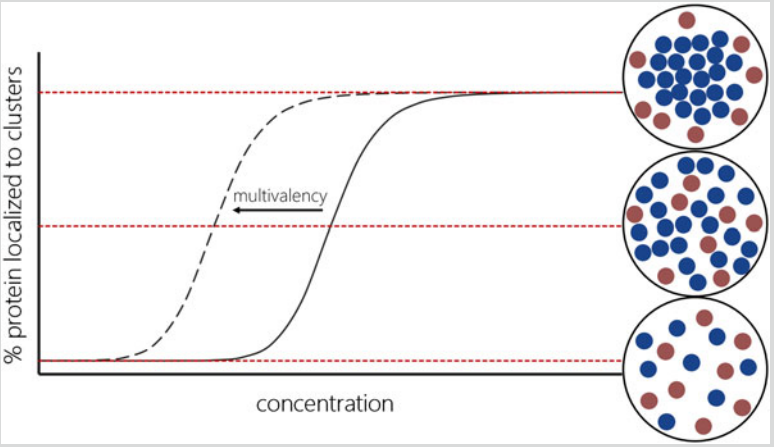
3.1 Multivalency

A number of recent experimental studies have highlighted how multivalent macromolecules, including both multi-domain proteins and RNA, are critical for the assembly of higher-order structures. Recently, the Rosen group systematically explored the importance of non-covalent interactions between multi-domain proteins in the formation of μm-sized cellular structures [55, 56]. Specifically, they studied the interactions between SRC homology 3 (SH3) domains and proline-rich motif (PRM) ligands, both of which are commonly found in cell signaling pathways. Chimeric proteins engineered with one to five SH3 domains were mixed incrementally with proteins containing up to 5 repeats of PRM ligand [56]. They found that a mixture of protein with four SH3 domains and four tandem PRM ligands resulted in an opalescent solution containing numerous droplets of 1–50 μm in diameter phase-separated from the bulk of the solution [56]. In contrast, no such structures could be formed in a solution of SH3 domains or PRM ligands alone. Proteins within the phase-separated droplets were concentrated by about 100-fold

relative to the surrounding solution. This phenomenon, commonly referred to as liquid–liquid demixing, is reminiscent of non-membranous structures *in vivo*, which maintain high local protein concentrations while allowing for fluid component exchange with the surrounding bulk phase. As indicated by this study, the transition point for forming such microscopically visible structures is usually sharp and dependent on molecule valency [55, 56]. Consistently, studies have suggested that a critical threshold for protein concentration is required to form RNA granules and that this amount is inversely correlated with the valency of the molecule (reviewed in [57, 58]; see Box 11.1).

Box 11.1: Phase Separation in Biology

A phase is a spatial region with uniform physical properties. In cells, discrete phases can form via a process of liquid-liquid demixing, resulting in droplet-like structures that remain distinct from the cytosol. This physical phenomenon has recently emerged in connection to non-membranous structures as a paradigm to understand cellular compartmentalization. Just as oil and water separate due to hydrophobic effects, proteins and nucleic acids can demix into condensed droplets above a certain concentration threshold. Beyond this transition point, multivalent and low affinity interactions mediated by PrLDs/LCDs create stronger intermolecular interactions within the droplet than in the surrounding cytoplasm. The components that localize to these structures therefore exist along the boundary between solubility and solid-like aggregation because their components bind transiently without entering a more ordered state. In other words, phase separation occurs when low-affinity interactions bring molecules together without arresting their dynamics [59].



(continued)

Box 11.1 (continued)
A graphical representation of the dependence of phase separation on protein concentration. Multiple studies have indicated that the transition from solubility to a phase-separated state is sharp [55, 60, 61] and multivalent interactions have been shown to lower this transition threshold (see main text).

The criteria for liquid-liquid demixing in biological systems are poorly understood. Though "nucleators" of some RNA granules have been identified—stalled initiation factors in SG assembly, for example—how these proteins initiate demixing while other complexes do not remains uncertain. Nevertheless, conceptualizing RNA granules as μm-sized cellular structures with unique biophysical properties allows us to gain insight into functions that might not be as evident from single-protein approaches, especially given the staggering complexity of RNA granule composition.

What might be the biological advantage of such non-membranous structures? First, since demixing itself is strongly concentration-dependent, a point of localized protein enrichment could regulate the rate of biochemical reactions, accelerating some by acting as "micro reactors" while sequestering away components to slow others [59]. Incorporating only the proteins with an affinity for LCD-mediated interactions might provide RNA granules with a passive selectivity mechanism, aided by multivalent proteins that contribute a "branching" scaffold to the assembly process. Furthermore, by virtue of their reversibility, which is mediated in large part by protein modifications, phase-separated structures can assemble only when needed, allowing for their composition to vary based on cellular requirements. For a more comprehensive discussion of liquid-liquid phase separations in biology, interested readers are directed to the recent review by Hyman, Weber and Jülicher [62].

The assembly of these phase-separated higher-order structures is critical for the physiological functions of the proteins that are recruited to them. For example, SH3–PRM interactions are observed naturally in the nephrin–NCK–N-WASP system—a three-component interaction on the membranes of kidney podocytes that is involved in the formation of the glomerular filtration barrier, which participates in actin assembly. Applying the same methodology as described above, the Rosen group found that interactions between the three SH3 domains of NCK and the six PRMs in N-WASP were sufficient to induce the formation of phase-separated droplets and stimulate actin assembly *in vitro*. In addition to three SH3 domains, each NCK molecule has one SH2 domain that can bind to phosphorylated tyrosine. Addition of the phosphorylated nephrin tail, which contains three phosphorylated tyrosine sites, significantly reduced the critical amount of NCK and WASP proteins required to form phase-separated droplets [56]. Thus, in cells, protein kinases can regulate the level of nephrin phosphorylation, which in turn regulates a cascade of multivalent interactions to form higher-order structures.

In addition to multi-domain proteins, phase transitions can be triggered by other macromolecules capable of multivalent interactions, including RNA—the categorical component of RNA granules. A single RNA can serve as a scaffold of repeating units that allows for the binding of multiple RNA-binding proteins. Parallel to the multi-domain protein studies, the Rosen group has also demonstrated that phase-separated droplets can be formed *in vitro* by incubating a tetravalent RNA-binding protein polypyrimidine-tract-binding protein (PTB) with an RNA oligonucleotide containing five repeats of the synthetic polypyrimidine tract, UCUCUAAAAA [56]. Analogous pathways of protein concentration via nucleating RNAs have been proposed for the higher-order assembly of FUS protein and nuclear bodies [61, 63, 64]. For example, Schwartz and colleagues [61] have recently shown the importance of RNA for FUS to assemble and bind to the C-terminal domain of RNA polymerase II. The addition of less-than-stoichiometric amounts of RNA to a FUS protein solution induced liquid droplets much more readily than occurred in a solution of FUS alone [61]. Combined with the finding that FUS binds RNA with high cooperativity (due to multiple RNA-binding sites), the authors suggest that FUS oligomerization is nucleated by a tightly packed initiating FUS–RNA complex [61]. Last but not least, it is worth noting that other cellular macromolecules besides proteins and RNA—such as poly-ubiquitin or poly(ADP-ribose)—are also found in SGs and could potentially serve as multivalent scaffolds by recruiting proteins to their repeating units [48, 65, 66] (see Sect. 11.3.3).

3.2 Intrinsically Disordered Regions

Intrinsic disorder has emerged as an essential mechanism in a diverse array of pathways, including the formation of RNA granules. Previously believed to be passive separators of functional domains, intrinsically disordered regions (IDRs) dynamically transition between structural states in order to serve multiple purposes in an efficient and highly regulated manner. A subclass of IDRs that is both essential to RNA granule physiology and holds important implications for neurological disease is the prion-like domain (PrLD). Prions are self-templating proteins originally discovered in yeast that serve in a variety endogenous functions and are capable of entering highly stable aggregates termed amyloids [67]. These characteristics stem from a prion domain of low amino acid complexity, typically 60 residues in length and enriched in asparagine, glutamine, tyrosine, and/or glycine residues [68]. Application of an experimentally validated bioinformatics algorithm [69] to the human proteome revealed that 1 % of proteins harbor predicted domains exhibiting a similar aggregation propensity to yeast prions [70]. Within this 1 %, however, there is a 12-fold enrichment for proteins that also contain at least one RNA recognition motif (RRM) [68]. Significantly, many of these "prionogenic" RBPs are known components of SGs [53] or RNA granules in general [60].

Studies from the McKnight group have elucidated possible roles for low complexity domains (LCDs) in the formation of RNA granules. McKnight and

colleagues fortuitously discovered that a small molecule chemical, biotinylated isoxazole (b-isox), precipitated a class of proteins from cell-free lysate exhibiting remarkable overlap with known components of RNA granules [60]. Despite an exceptionally high enrichment for RBPs within this precipitated class, the LCDs, not the RRMs, of these RBPs were found to be necessary and sufficient for b-isox mediated precipitation out of lysate [60]. Notably, purified samples of the FUS or hnRNPA2 LCDs formed hydrogels at high concentrations via oligomerization-mediated phase transition [60]. Besides this capacity for self-association (homotypic trapping), FUS and hnRNPA2 LCD-containing hydrogels were also able to retain the LCDs of other RBPs [60]. As revealed by transmission electron microscopy and light microscopy, the hydrogels are composed of morphologically uniform amyloid-like fibers which incorporate other proteins via LCD interactions [60] (see "Notes"). However, unlike pathogenic fibers in amyloid plaques and prion diseases, the formation of these fibers is reversible. Conjoint to these studies, the McKnight group also analyzed the mRNAs pulled down by b-isox mediated precipitation using high-throughput sequencing [71]. They found that the precipitated mRNAs are also closely aligned with those known to partition into RNA granules. More intriguingly, the authors found that these mRNAs could also be retained by hydrogels made by the FUS LCD alone. This surprising selective mRNA retention suggests a model wherein RBPs associate with one another via their LCDs, pulling in select mRNAs via an RNA-binding domain [71].

The utility of PrLDs/LCDs is grounded in their ability to self-associate, and this propensity to rapidly oligomerize through low affinity protein–protein interactions might underpin the dynamic physical properties that characterize SGs and other RNA granules [32, 53]. Consistent with how readily SGs have been observed to exchange components and dissolve after stress release, an additional advantage for PrLD/LCD-mediated aggregation might therefore be high reversibility. The regulatory mechanisms that control these transient interactions are also being explored and will be discussed in subsequent sections. We note, finally, that PrLDs are typically also composed of low complexity amino acid sequences, frequently Q/N or QGSY-rich, and that both are subtypes of intrinsically disordered domains. The two terms PrLD and LCDs are often used interchangeably in the literature and are synonymous in this text as well.

3.3 SG Assembly

As with their function, the precise mechanisms of SG assembly are not fully understood and likely vary alongside composition under different stress states or points of disruption to the translation initiation processes [7, 23]. Since SGs form in response to translational arrest and almost always contain components of the translation initiation complex, the assembly process likely begins with a pool of repressed, translationally inactive mRNPs. Pathways of assembly that proceed from this starting

pool of aborted complexes likely involve an mRNP-specific series of RNA–protein and protein–protein interactions mediated by PrLDs/LCDs. Although the initial stages of this process are not understood, there are some known examples of SG nucleation by oligomerization and multivalency-mediated cross linking. The SG component TIA-1, for example, contains a C-terminal PrLD enriched in glutamine residues, bearing high similarity to the aggregation-prone NM domain found in yeast prion Sup-35 [72]. In humans, both the prion-like and RNA-binding domains of TIA-1 are necessary for SG assembly [32]. While the PrLD confers a propensity for aggregation by homotypic oligomerization, the RNA-binding domain is required to recruit RNAs to the complex [32]. Indeed, without the RRM, recombinant TIA-1 PrLD fragments form constitutive micro-aggregates when overexpressed, sequestering endogenous TIA-1 and inhibiting SG formation [32]. G3BP1, another common SG marker, might mediate assembly in a similar fashion. G3BP1 harbors both an NTF2-like domain prone to dimerization as well as an RNA-binding domain [73]. Though not a PrLD, the NTF2-like domain in conjunction with RNA-binding imbues G3BP1 with the ability to interact with multiple partners. As discussed in Sects. 11.4 and 11.5, multiple RBPs which localize to SGs, such as TDP-43 and hnRNPA2, also contain LCDs. Pathological mutations in these regions might be responsible for impairing the reversibility of RNA granule formation in neurons, resulting in neurodegeneration. Thus, the preponderance of SG components with both LCDs and RBDs, the in vitro observation of liquid/liquid demixing facilitated by these properties, and the capacity for some of these components to trigger SG formation upon overexpression builds a strong case that reversible protein/RNA crosslinking plays a role in the assembly process. Importantly, pathways for how these mechanisms might be regulated are also beginning to be uncovered. Chaperone proteins and post translational modifications both play major roles in controlling the selective aggregation of heterogenous mRNPs into functional, higher order structures.

3.3.1 Regulation by Chaperones

Prion-like interactions are a double-edged sword. They provide many of the key features of RNA granule physiology, but are prone to non-specific, potentially pathological aggregation as well. Recent data suggest that protein chaperones are required to continuously keep PrLDs in check, titrating the number of disordered domains available to interact in the cytoplasm, or altering their conformations to inhibit aggregation. In yeast, for example, highly stable amyloids can be returned to solubility by the heat shock protein HSP104, which works in conjunction with HSP40 and HSP70 to refold and reactivate denatured proteins locked within aggregates, thereby acting as a disaggregase [74]. Indeed, the formation and propagation of TIA-1 aggregates in yeast is entirely dependent on the activity of HSP104, a known regulator of many Q/N-rich proteins [75]. Interestingly, HSP70 is expressed at higher levels after transfection with the PrLD of TIA-1 and can inhibit the latter's aggregation when

cotransfected into COS7 cells [32]. It is also known that under sodium arsenite treatment, HSP70 mRNA is shielded from recruitment into SGs by the translation regulating protein YB-1, resulting in its preferential translation under stress [33]. Besides YB-1, an additional mechanism of control has been elucidated by Boyault et al. [76], based on the activation of heat-shock factor 1 (HSF1) by the cytoplasmic deacetylase HDAC6, which has known disaggregase function as well. Under basal conditions, HSF1 resides in a dormant complex along with HDAC6 and HSP90. Upon binding ubiquitin, HDAC6 releases HSF1 to transcriptionally activate a number of heat shock protein-encoding genes, including HSP70 [76]. SGs were found to be enriched for ubiquitin and HDAC6 was shown to be an indispensable component under a wide variety of stress conditions [48]. Fascinatingly, the HDAC6-mediated activation of HSF1 requires another disaggregase—the chaperone-like AAA ATPase P97/VCP, which is itself a SG component—in order to "recycle" ubiquitinated HDAC6 back into the dormant HDAC6/HSF1/HSP90 complex [76]. As will be discussed later, mutations in the gene encoding VCP have been implicated in the neurodegenerative disease ALS. Thus, heat shock proteins may be a general protective mechanism in the regulation of unfolded proteins, either lowering the concentration of prion-like substrates or inducing conformational changes that make them less aggregation-prone [77]. In both cases, it seems likely that proteins containing PrLDs/LCDs are continuously kept in check by heat shock proteins, and that a shift in chaperone concentration can induce aggregation [32].

3.3.2 Regulation by Post Translational Modifications

Finally, post-translational modifications have also been observed to enhance or inhibit assembly of SGs. In the case of G3BP1, oligomerization is blocked by phosphorylation on Ser149, and dephosphorylation is required for its ability to nucleate SGs [73]. Similarly, hnRNPA1 is only recruited to SGs when phosphorylated by Mnk1/2 [78], whereas TTP and BRF1, two common SG components that bind to AU-rich element (ARE)-containing mRNAs, can be excluded by MK2 and PKB mediated phosphorylation, respectively, with downstream effects on ARE transcript degradation [31, 79]. Phosphorylation of human antigen R (HuR) by Janus kinase 3 (JAK3) prevents its recruitment into SGs, accelerating the decay of SIRT1 and VHL mRNAs [80]. A recent study by the Pelkmans group has demonstrated a role for phosphorylation of LCDs in the regulation of SG assembly/disassembly. Dual specificity kinase DYRK3 cyclically partitions between SGs and the cytosol [51], activity which has been proposed to be regulated by auto-phosphorylation at the N-terminal LCD. Consistent with this proposal, the expression of N-terminal LCDs alone or a kinase-dead mutant of DYRK3 induces SG formation in the absence of stress, while inhibition of this kinase by GSK-626616 prevents SG disassembly [31, 51]. Thus, phosphorylation functions as a protein-specific and reversible mechanism to either prevent or trigger recruitment of components to SGs.

Besides phosphorylation, poly(ADP-ribosyl)ation has been implicated in regulating SG assembly/disassembly [66]. Poly(ADP-ribosyl)ation refers to the addition

of two or more ADP-ribose units onto proteins, where the addition is regulated by a family of 17 ADP-ribosyltransferases, commonly referred to as PARPs [81]. SGs contain five out of the 17 PARPs as well as two isoforms of the degradative poly(ADP-ribose) glycohydrolase PARG in humans [66]. Three pieces of evidence suggest that poly(ADP-ribosyl)ation is intrinsically linked to SG structural integrity [66]. First, overexpression of any of the SG-localized PARPs induces SG formation in the absence of stress. Second, overexpression of PARG inhibits SG formation in the presence of stress. Third, PARG knockdown delays SG disassembly. Several proteins, including G3BP1 and TIA-1, are poly(ADP-ribosyl)ated upon stress, but the role of these modifications in SG assembly and function remains to be determined. A recent informatics analysis indicate that RNA granule proteins enriched for LCDs are preferentially poly(ADP-ribosyl)ated, indicating poly(ADP-ribose) may potentially direct cellular organization [65].

To summarize, multivalent protein–protein and RNA–protein interactions underlie the formation of RNA granules, which resemble liquid droplets phase-separated from the bulk of solution *in vitro*. The fluid-fluid demixing observed in the formation of these granules relies on a protein species' ability to oligomerize, forming a scaffold and incorporating other protein or RNA components as binding partners [58]. Emerging research increasingly points towards the role of intrinsically disordered regions, such as PrLDs/LCDs, as necessary in the assembly of SGs and other RNA granules. These regions likely switch conformational states to facilitate/disengage interactions within RNA granules, which are regulated by a combination of post-translational modifications and chaperones [82]. As a whole, the system relies on a delicate equilibrium between intrinsic disorder and the regulatory mechanisms that restrain misaggregation. However, the complexity of this balancing act translates to openings for disruption, and SGs specifically have been inextricably linked to the pathogenesis of two neurological diseases in particular—ALS and FTLD.

4 ALS and FTLD: When Aggregation Goes Awry

At the present time, ALS and FTLD are among the best-characterized neurological diseases, and both increasingly appear to be at least partially linked to endogenous RNA granules. ALS is a clinical entity characterized by the progressive loss of motor neuron function, resulting in muscle wasting and eventually respiratory failure [83]. Approximately 10 % of cases have familial origin (f.ALS), with the remaining majority occurring in individuals sporadically (s.ALS) [83]. In contrast, frontotemporal lobar degeneration (FTLD), the second most common form of dementia behind Alzheimer's, proceeds with neuronal atrophy in the frontal and temporal cortices, causing patients to undergo changes to personality, behavior, and language abilities [84]. Both diseases are accompanied by a wide variety of possible genetic mutations resulting in clinical heterogeneity and notably, symptomatic overlap in a large subset of patients, suggesting that the two are linked by a spectrum of pathologies [85]. This section will overview one facet of the many

genetic and molecular factors that comprise the ALS/FTLD disease system: the pathological role of RBPs and RNA metabolism. Special emphasis will be paid to the mechanisms by which many of the same protein properties that allow for proper SG function can be co-opted or otherwise disrupted by genetic and environmental perturbations in these neurological diseases. Several hypothesized pathways of neurodegenerative toxicity are subsequently discussed.

4.1 SGs and Cytoplasmic Pathological Inclusions

One of the few features that is common across the majority of ALS and FTLD variants is the presence of abnormal cytoplasmic inclusions in patient histology. The morphology, composition, and tissue distribution of these inclusions can vary but they nevertheless bear a resemblance to physiological non-membranous granules. It was the discovery that TAR DNA-binding Protein 43 (TDP-43) and Fused in Sarcoma (FUS) were components of the majority of these inclusions that both shifted attention to the role RBPs might play in neurodegeneration and suggested a connection to SGs in neurological pathogenesis [86, 87].

TDP-43 is a member of the hnRNP family and contains two putative RRMs biased towards GU-rich intronic targets [88] or exceptionally long introns [89]. In addition, it has a glycine-rich C-terminal PrLD and a bipartite nuclear localization signal (NLS) between the N-terminus and RRM1 (Fig. 11.2). Following the revelation that TDP-43 was a component of pathological inclusions in ALS/FTLD models, the first concrete link to SGs came from observations that TDP-43 and some of its pathological mutants also co-localized with canonical SG markers in a diverse array of cell types, stress conditions, and mutation models. Indeed, though mutations to TARDBP, the gene encoding TDP-43, are implicated in only 4 % of f.ALS cases and are rarer still in s.ALS [95], TDP-43 protein is present in 45 % of FTLD and a full 97 % of ALS inclusions [85]. It is now known that endogenous TDP-43 is a consistent component of SGs [86]. HeLa cells subjected to oxidative or heat stress, for example, exhibit up to 56 % depletion of TDP-43 from the cytoplasm, corresponding with a quantitative increase in RIPA buffer-insoluble TDP-43 localized within putative SG aggregates [90]. Though TDP-43 knockdown does not in itself prevent SG assembly, lowered expression of TDP-43 has been shown to alter cytoplasmic levels of two other SG components, TIA-1 and G3BP1, causing a delay in SG formation, the formation of smaller SGs, lowered persistence after stress release, and diminished cell viability [90]. A follow-up report further confirmed that the modulation of G3BP1 mRNA levels by TDP-43 dictates SG size [96]. Most recently, the Vande-Velde group showed that by indirectly controlling the G3BP1-dependent coalescence of smaller SGs into larger structures, TDP-43 also affects SG–PB docking, which partially determines mRNA fate after interruptions to translation [97]. siRNA against TDP-43 resulted in the progressive and step-wise loss of poly(A)mRNA after oxidative stress, suggesting a functional role in directing the flow of mRNA transcripts out

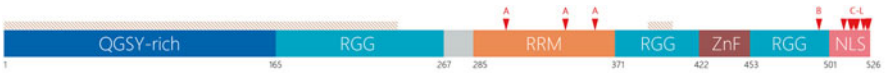
	Mutation	Pathogenic model: Disease	Pathogenic model: Cell line	Colocalization with stress granules: Stress	Colocalization with stress granules: Colocalization?	Colocalization with stress granules: Marker	Ref.
A	K83A, R84A, K85A	engineered	HeLa	HS, SA, CL	Y	TIA-1	93
	K83A, R84A, K85A	engineered	primary rat hippocampal	HS	Y	TIA-1	93
B	D169G	ALS	patient lymphoblasts	SA	N	TIA-1	90
C	G294A	ALS	BE-M17	SA	Y	TIA-1	91
D	A315T	ALS	HeLa	HS	N	TIA-1	93
	A315T	ALS	BE-M17	SA	Y	TIA-1	91
	A315T	ALS	primary rat hippocampal	overexpression	Y	TIA-1, G3BP	94
E	Q331K	ALS	BE-M17	SA	Y	TIA-1	91
F	M337V	ALS	HeLa	HS	N	TIA-1	93
G	Q343R	ALS	primary rat hippocampal	overexpression	Y	TIA-1, G3BP	94
	Q343R	ALS	BE-M17	SA	Y	TIA-1	91
H	G348C	ALS	HeLa	HS	N	TIA-1	93
	G348C	ALS	stable HEK293T	Sorbitol	Y	hnRNPA1, HuR, TIAR	92
I	R361S	ALS	patient lymphoblasts	SA	N	TIA-1	90

Fig. 11.2 Schematic representation of the functional domains of TDP-43. TDP-43 is a 414-amino acid RBP with functions in transcriptional repression, splicing, and translational regulation. The vast majority of ALS-associated mutations occur within the C-terminal glycine rich PrLD. Table only includes data on studied links between ALS mutations and SG association in cellular models. Mutations are marked by red arrows and letters corresponding to table entries. *NLS* nuclear localization signal, *RRM* RNA recognition motif, *NES* nuclear export signal, *Tan shading* PrLD [90–94]. *HS* heat shock, *SA* sodium arsenite, *CL* clotrimazole

of stalled polysomes. In the context of the RNA triage model, the fact that poly(A) levels are diminished despite the fact that SG–PB docking is also attenuated can only mean that transcripts are routed directly to PBs for degradation in the absence of G3BP1 [97]. Therefore, by regulating levels of TIA-1 and G3BP1, TDP-43 contributes to both SG assembly and mRNA triage.

FUS carries a contiguous pair of PrLDs at the N-terminus, an RRM, two RGG domains flanking a zinc finger, and a NLS at the C-terminus [98] (Fig. 11.3). FUS, though significantly less abundant in neurological inclusions (<1 % in ALS, ~9 % in FTLD), shares many functional similarities with TDP-43—they both shuttle between the cytoplasm and nucleus, play roles in RNA metabolism, and exhibit prion-like behavior [85]. Unlike TDP-43, SGs form and behave normally in FUS-depleted cells [96]. Furthermore, endogenous FUS is not a typical SG component, remaining predominantly nuclear during the stress response [96, 99, 100], though there have been observations of some WT FUS association with the SG marker TIA-1 in a fraction of stressed cells [106, 107], and more particularly in cells subjected to "hyperosmolar" stress [96, 108]. Interestingly, however, siFUS-treated cells have been shown to undergo a threefold increase in the number of P-bodies labeled with Dcp1a, indicating a partial role in mRNA triage [97].

Strikingly, in ALS and FTLD disease models, recruitment of both proteins to SGs is frequently observed and can be dysregulated by the introduction of mutations. Multiple known SG components have been shown to localize with cytoplasmic inclusions seen in ALS/FTLD patients. For example, overexpression of wild-type TDP-43 or transfection with truncated TDP-43 fragments commonly



	Mutation	Pathogenic model: Disease	Pathogenic model: Cell line	Colocalization with stress granules: Stress	Colocalization?	Colocalization with stress granules: Marker	Ref.
A	F305L, F341L, F359L	engineered	N2a	SA	N	TIAR	104
	R495X	ALS	HEK-293	SA	Y	TIAR	101
B	R495X	ALS	Zebra fish embryonic spinal neurons	HS	Y	TIAR	101
	R495X	ALS	HEK-293	overexpression	Y	PABP1, TIA-1	100
C	K510X	engineered	CV-1, rat primary neurons	overexpression, SA*	Y	PABP, TIA-1	103
	R514S	ALS	HeLa	overexpression	Y	TIA-1	102
D	R514G	ALS	CV-1, rat primary neurons	overexpression, SA*	Y	PABP, TIA-1	103
	R514G	ALS	patient fibroblasts	SA*	Y	PABP1	103
E	R514S, R521C, P525L	engineered	HeLa	overexpression	Y	G3BP	102
F	G515X	ALS	HEK-293	SA	Y	TIAR	101
	G515X	ALS	Zebra fish embryonic spinal neurons	HS	Y	TIAR	101
G	P517L	ALS	N2a	overexpression	Y	eIF3	105
H	H517Q	ALS	Zebra fish embryonic spinal neurons	HS	N	TIAR	101
	R518K	ALS	N2a	SA	Y	TIAR	104
	R521C	ALS	patient fibroblasts	SA*	Y	PABP1	103
	R521G	ALS	HeLa, primary neurons	HS	Y	PABP1	99
I	R521G	ALS	HEK-293	SA	Y	TIAR	101
	R521G	ALS	Zebra fish embryonic spinal neurons	HS	Y	TIAR	101
	R521G	ALS	HEK-293	overexpression	Y	PABP1, TIA-1	100
J	R522G	ALS	HeLa, primary neurons	HS	Y	PABP1	99
K	R524S	ALS	HeLa, primary neurons	HS	Y	PABP1	99
	P525L	ALS	HeLa	overexpression	Y	TIA-1	102
L	P525L	ALS	HeLa, hippocampal neurons, SH-SY5Y neuroblastoma cells	HS, SA, CL	Y	TIA-1	93
	P525L	ALS	HeLa, primary neurons	HS	Y	PABP1	99

Fig. 11.3 Schematic representation of the functional domains of FUS. FUS is a 526-amino acid RBP with functions in transcriptional repression, DNA damage repair, and splicing. The majority of ALS-linked FUS mutations occur within its NLS, linking its mislocalization to neurodegenerative pathogenesis. Table only includes data on studied links between ALS mutations and SG association in cellular models. Mutations are marked by red arrows and letters corresponding to table entries. *NLS* nuclear localization signal, *RRM* RNA recognition motif, *ZnF* zinc finger domain, *RGG* arginine-Glycine-Glycine-rich domain, *Tan shading* PrLD [93, 99–105]. *HS* heat shock, *SA* sodium arsenite, *CL* clotrimazole, (*Asterisk*) co-transfection of mutant FUS with the wild-type resulted in the recruitment of both to SGs

observed in patients exhibiting TDP-43 pathology yielded co-localization with SG markers in human neuroblastoma BE-M17 cells even in the absence of stress [91]. In the same study, TDP-43+ inclusions were shown to overlap with eIF3 and TIA-1 in both ALS and FTLD donor brain samples [91]. These parallels between cytoplasmic inclusions and SGs have spurred considerable interest in how endogenous mechanisms of RNA granule regulation can go awry, resulting in the altered composition, stability, and accumulation that categorically distinguish pathogenic inclusions from those observed in the physiological stress response.

4.2 FUS, TDP-43 and Neurodegenerative Pathogenesis

The multimerization-mediated phase transitions that imbue RNA granules with their dynamic, fluid-like properties rely on the tight regulation of protein concentration (see Box 11.1). A considerable body of evidence now indicates that disruptions

to the equilibrium of protein concentration are causative factors in neurodegenerative pathogenesis. In both ALS and FTLD, a striking number of known mutations disrupt TDP-43 and FUS shuttling between the nucleus and cytoplasm or enhance their propensity to aggregate, with implications in the formation of insoluble cytoplasmic inclusions and ultimately cell toxicity.

4.2.1 Mislocalization of FUS from the Nucleus to the Cytoplasm

A number of C-terminal missense mutations linked to ALS have been shown to disrupt nuclear import of FUS, resulting in cytoplasmic accumulation [99]. Importantly, the degree to which nuclear import is disrupted seems to correlate with the severity of the disease. Mutations which result in higher levels of cytoplasmic FUS accumulation were identified in pedigrees exhibiting earlier age of onset [99, 101]. This trend may be partially explained by the crystal structure of the FUS NLS when bound to its nuclear receptor, Transportin. Unlike the localization signals of other hnRNPs which bind Transportin in a fully extended conformation, the FUS NLS forms an α-helix structure when bound [109]. This α-helix is flanked by an N-terminal hydrophobic motif and a series of basic residues, both of which are conserved in the FUS-related proteins EWS and TAF15 [109]. ALS-associated mutations are clustered within these two flanking epitopes, and a particularly severe mutation identified in early-onset juvenile ALS (P525L)—a residue that makes numerous contacts with Transportin—results in a ninefold loss in binding affinity [109].

Dormann and colleagues [110] subsequently found that the third RGG domain (RGG3) of FUS can also facilitate binding with Transportin, even if the NLS is disrupted by the P525L mutation. However, this FUS RGG3–Transportin interaction can be negated by methylation. This observation led to the surprising finding that cytoplasmic FUS/SG aggregates induced in cells by transient transfection of the P525L mutant were enriched for methylated FUS [110]. Post-mortem tissue samples from ALS-FUS patients exhibiting the same cytoplasmic aggregates also contained methylated FUS, whereas three FTLD-FUS subtypes did not [110], possibly suggesting that the pathogenesis of these two closely related diseases may differ. Taken together, Dormann and colleagues posit that epitopes in the NLS anchor FUS to the Transportin binding pocket, allowing RGG3 to stabilize the interaction. Only when the NLS and RGG3 domains are both disrupted, by mutations and methylation, respectively, is nuclear import prevented [110].

4.2.2 Prionogenicity of FUS and TDP-43

Whereas endogenous SGs are reversible and dynamic by nature, stability and persistence are a token trait of pathological inclusions in patient histology [102]. One particularly compelling explanation for this physical misbehavior has implicated the PrLDs/LCDs present in both FUS and TDP-43 as susceptible to mutation-induced fibrilization [68]. Indeed, the endogenous forms of both proteins are known to be

aggregation-prone, coalescing into phase-separated droplets [111], or a polymer hydrogel [60] when purified *in vitro*. In bioinformatics assays, based on the predictive algorithm developed by Alberti and colleagues [69], both FUS and TDP-43 exhibit an exceptionally high prion propensity, ranking 15th and 69th, respectively, out of 27,879 analyzed proteins [69, 112]. The predicted FUS PrLD encompasses amino acids 1-239, spanning the entirety of its QGSY-rich N-terminus, and most of the adjacent glycine-rich domain. The LCD used by Kato and colleagues to generate a hydrogel *in vitro* covers residues 2-214 (See Sect. 11.3.2) [60]. In the case of TDP-43, a PrLD spans 138 residues at the C-terminus [112]. Emergent data now suggests that mislocalization allows these physical properties to come to bear in the cytoplasm, resulting in aggregation and association with SGs. ALS-linked mutations exacerbate this behavior by altering the PrLD itself in the case of TDP-43, or by elevating cytoplasmic FUS concentration.

4.2.3 Aggregation-Prone Mutations in FUS and TDP-43

Current models indicate that FUS aggregation is driven by its mislocalization, facilitated by mutations to its NLS and exacerbated by stress-induced recruitment into SGs [99, 101]. For example, HEK293 cells stably expressing FUS truncation mutants R495X or G515X at near endogenous levels exhibited stable FUS aggregates, corresponding with a cytoplasmic:nuclear FUS expression ratio 30–50 times greater than in cells expressing the wild-type [101]. Unlike TDP-43, the first RGG domain of FUS is required in addition to the PrLD in order to aggregate [113, 114]. Interestingly, this RGG domain houses a short second sequence of prion-like residues which narrowly misses the cutoff of Alberti's prediction algorithm [114]. In addition, aggregation of FUS with SGs is also RNA-dependent as demonstrated by the fact that FUS association with canonical SG markers can be abolished by RNAse treatment [103]. Finally, although FUS localization to the cytoplasm is a prerequisite for aggregation with SGs, mutant FUS has been shown to trap endogenous (presumably nuclear) FUS into SG aggregates, most likely by protein–protein interactions facilitated by the FUS PrLD [103].

TDP-43 appears to form inclusions by a different sequence of events, perhaps due to its endogenous cytoplasmic function during the stress response. Specifically, the vast majority of TDP-43 mutations are clustered in its C-terminal PrLD, and an emerging model is that these mutations prevent its return to the cytoplasm post stress release by strengthening the protein–protein interactions that drive its incorporation into SGs [115]. As an example, ALS-associated mutations have been found to promote TDP-43 aggregate stability [116]. Specifically, radioactive pulse labeling showed that TDP-43 mutants are degraded as much as 4 times slower than wild-type [116]. Furthermore, two PrLD mutations in TDP-43—Q331K and M337V—were found to significantly increase association between TDP-43 and FUS in the nucleus, an interaction that is not normally observed and not present in patient histology [116]. These same two mutations consistently form more cytoplasmic puncta in yeast models and aggregate more quickly *in vitro*, demonstrating that point mutations can enhance PrLD-mediated interactions [111]. Though this study did not

examine co-localization with SG markers, mutations in the TDP-43 PrLD have also been found to alter the dynamics of SG assembly under sorbitol treatment, an oxidative and osmotic stressor capable of targeting endogenous TDP-43 to cytoplasmic structures enriched for the canonical SG markers TIAR and HuR [92]. Under prolonged sorbitol exposure, a G348C mutation was found to increase the fraction of primary glial cells containing TDP-43+ SGs, accompanied by the formation of fewer, larger SGs compared to cells expressing the wild-type [92].

4.2.4 Mislocalization Without Aggregation

An important caveat is that in many cellular models of FUS or TDP-43 pathogenesis, cytoplasmic relocalization is not sufficient to induce aggregation. Absent cellular stress, FUS bearing ALS-associated NLS mutations, for example, was not observed to aggregate or associate with SGs despite extensive cytoplasmic localization [99]. Indeed, in most cellular models, even if aggregates do ensue from the mislocalization of FUS alone, they are typically only detected in a minority of cells [101] and exposure to stress increases both the number of cells with cytoplasmic aggregates and the average number of inclusions per cell [101]. Dorman and colleagues propose a "2-hit" model for pathogenicity, whereby mutations to the C-terminal NLS first causes FUS to mislocalize to the cytoplasm (step 1), where it then becomes susceptible to stress-induced aggregation and recruitment into SGs (step 2). Similarly, the TDP-43 mutants observed to alter SG assembly dynamics remained nuclear until prompted to relocalize by an external stressor [92]. In a separate study, three ALS-associated TDP-43 mutations, all occurring in the C-terminal PrLD, were confined to the nucleus when transiently transfected into HeLa cells and therefore did not recruit to SGs upon heat shock [93]. Consistent with the working "2-hit" model of TDP-43/FUS proteinopathy, all three of these mutations, when introduced in conjunction with a disrupted NLS, resulted in cytoplasmic localization, aggregation, and SG recruitment [93].

4.3 Aggregation, Toxicity, and a Role for SGs

Explicit clarification is warranted here to elucidate the relation of cytoplasmic aggregation to cellular toxicity. At the present time, the data for whether aggregation is necessary to convey cellular toxicity is mixed, and how pathogenesis proceeds from cytoplasmic accumulation of TDP-43 or FUS is unknown. Observations from cellular models that toxicity is accompanied by cytoplasmic aggregation are legion, but many reports question whether inclusion formation plays a causative role in pathogenesis or is instead an endogenous cellular response to other disease factors. Gitler and colleagues evaluated molecular determinants for both TDP-43 and FUS mediated toxicity in a yeast model recapitulating the cytoplasmic aggregation observed in ALS patient histology [114, 117]. While the C-terminal PrLD of TDP-43 was found to be necessary and sufficient for both cytoplasmic localization and aggregation,

toxicity required an intact RRM2 domain as well [117]. Interestingly, constructs bearing only the two RRMs or the C-terminal PrLD were observed to localize and aggregate in the cytoplasm, despite the fact that neither was associated with abnormal cell growth [117]. FUS was found to depend on a more complex array of molecular determinants, requiring its N-terminal PrLD, RRM, and first RGG domain to aggregate in the cytoplasm and promote cell death [114]. Notably, as with TDP-43, this toxicity has been reported to depend on RNA-binding as deletion of the RRM rescues the deleterious effects associated with NLS mutants [104]. In a final example, Bosco et al. [101] observed that cells expressing R495X or G515X FUS mutants, both of which robustly localize to the cytoplasm and form aggregates, remained perfectly viable, and no signs of toxicity were identified. As will be discussed, these conflicting data on TDP-43/FUS toxicity might be partially explained by the intrinsic features that are unique to neurons, making the selection of proper cellular systems especially critical for studies examining pathways of toxicity.

4.3.1 Neuron-Specific Toxicity

Some have argued that cytoplasmic localization, not the formation of inclusion bodies, provides a superior correlate for cell death [115]. In support of this, preventing TDP-43 shuttling out of the nucleus by a mutation in the nuclear export signal rescued the toxicity of A315T mutants [115]. Studies arguing either that mutations in the PrLD do not cause cytoplasmic mislocalization until stress induction, or that mislocalization is not toxic—typically do not use a primary neuronal cell line, neglecting possible "neuron-specific mechanisms that govern metabolism and distribution" [115]. Indeed, the TDP-43 mutations A315T, G348C, and A382T, were only found to aggregate in the cytoplasm and confer toxicity when expressed in motor neurons derived from primary mouse spinal cord cultures, as opposed to tissue culture cell lines COS1 and N2A in which they remained confined to the nucleus and did not influence cell viability [118]. In Drosophila, ectopic localization of FUS was found to be crucial to development of a neurodegenerative phenotype and a simple deletion of the FUS nuclear export signal rescued toxicity [119]. One possible explanation is that nervous tissue is enriched for transcripts carrying longer introns compared to other tissues, a characteristic which might carry heightened susceptibility to FUS abnormalities [88, 120].

4.3.2 Stress and Inclusion Solubility

Another notable difference between in vitro studies and patient pathology is that the majority of cell studies aiming to model attributes of ALS and FTLD pathology do not recapitulate the irreversibility of TDP-43 and FUS inclusions in patient histology, with FUS or TDP-43 positive aggregates disassembling readily after removal of stress [93]. In a unique study utilizing mild but prolonged conditions of oxidative stress induced by paraquat in the neuron-like cell line SH-SY5Y, TDP-43 co-localized

with the SG markers HuR and TIAR in cytoplasmic inclusions [121]. Intriguingly however, these TDP-43 aggregates persisted after stress release or cycloheximide treatment despite the fact that HuR and TIAR had both returned to solubility [122]. These observations align with the developing notion that the PrLD-mediated assembly of RNA granules might "seed" the progression of aberrant protein aggregates by raising the local concentration of proteins bearing prion-like properties. Alternatively, a combination of post-translational modifications, chaperone-facilitated protein remodeling, and other mechanisms normally involved in RNA granule regulation may be disrupted under conditions of abnormal cellular stress. Indeed the paraquat-induced TDP-43 inclusions were heavily enriched for ubiquitin [122]. These phenomena might be unique to both cell and stress type, potentially explaining the conflicting findings of many in vitro studies of ALS/FTLD pathogenesis.

4.3.3 Disruptions to Endogenous TDP-43/FUS Function

If toxicity does not require aggregation, what might be the driving mechanism for TDP-43/FUS-mediated neurodegeneration? To date, TDP-43 and FUS functions have not been fully characterized but both proteins are known to play prolific roles in diverse pathways. Based on what is known, besides a common PrLD-mediated predisposition towards aggregation, FUS and TDP-43 do share some functional overlap. For example, the two proteins have been shown to complex in order to co-regulate the mRNA of HDAC6, the SG component with disaggregase properties and a primary role in activating HSF1 as discussed in Sect. 11.3.3 [123]. Moreover, TDP-43 mediates the alternative splicing of FUS mRNAs [124]. However, there is also considerable evidence for the divergence of TDP-43 and FUS-mediated pathogenesis. In addition to differences in domain requirements for toxicity, yeast screens have shown that there is very little overlap between genetic modifiers of TDP-43 and FUS toxicity [114]. Depletion of FUS alters the splicing of approximately 1000 mRNAs, most of which differ from those altered upon TDP-43 depletion [89]. Furthermore, TDP-43 and FUS have only very rarely been observed to co-localize in immunostaining of pathological inclusions in patient histology [114] and their binding partners share little overlap [89]. Thus, the root cause of pathogenesis might be the loss of TDP-43/FUS function, which can certainly be facilitated by cytoplasmic aggregation, but not necessarily so. In this model, aggregation induced by mutations, exogenous stressors, or some other as-of-yet unidentified pathway would be an easily visible symptom of disrupted TDP-43/FUS function. The direct cause of pathogenesis, however, would be the resulting dysregulation of RNA homeostasis, which would differ between TDP-43 and FUS pathologies. What is increasingly certain is that the expression of both TDP-43 and FUS is tightly regulated, and a variety of cellular and animal models have demonstrated the deleterious outcomes of altering expression levels of either.

Studies in C. elegans have recapitulated neurological impairment due to expression of human or mutant TDP-43. A transgenic line expressing human WT TDP-43 or a truncation mutant, for example, exhibited severe locomotor deficits, impaired

synaptic transmission, and growth defects despite an absence of detectable neuronal loss [35, 36]. In this case, though nuclear aggregates were detected in WT-expressing worms and cytoplasmic aggregates were observed in those expressing the truncation mutant, both sets of inclusions were assessed to be relatively immobile by FRAP [35, 36]. Importantly, RNAi of HSF1 was also found to dramatically worsen the neurological phenotype [35, 36]. In mammals, an important clue for how TDP-43 loss of function might proceed comes from transgenic mouse models in which mislocalization and aggregation are either rarely observed or not observed at all in post-mortem brain samples [125, 126]. In one study, expression of human TDP-43 was found to correspond with a loss of endogenous mouse TDP-43 in the nucleus—not to the cytoplasm, but through TDP-43's ability to autoregulate its own mRNA by non-sense mediated decay [126]. The fact that regions of the mouse brain that did not exhibit signs of neurodegeneration also did not show altered levels of endogenous mouse TDP-43 strongly indicates that the progression of the neurodegenerative phenotype is caused by TDP-43 depletion [126]. Genome-wide splicing arrays have also been used to show that depletion of endogenous mouse TDP-43 causes an enhancement of some splicing activities but an abrogation of others, while mice expressing a Q331K TDP-43 mutant exhibited mutation-unique splicing alterations [125]. Taken together, these studies demonstrate a strong dependence of TDP-43 function on levels of TDP-43 expression, which is not dependent on inclusion formation, but which does not exclude a role for cytoplasmic aggregation either.

In Vivo models for FUS pathologies also yield insight into the deleterious outcomes of altering its expression levels. Transgenic mice homozygous for an inactive FUS allele, for example, fail to suckle and die soon after birth [127]. Knock down of the FUS homologue Caz in Drosophila causes degeneration of motor neurons and abnormal locomotive behavior [128] whereas expression of wild-type or pathogenic mutations of human FUS in Drosophila induces a similarly severe neural impairment [129]. Therefore, the expression level of FUS protein needs to be critically maintained for normal neuronal function. As in the case for TDP-43, although cytoplasmic aggregates seem unnecessary for pathogenesis, the formation of such higher-order structures likely disrupts the delicate equilibrium of protein concentration required for the endogenous functions of these RBPs. In support of this, two studies have successfully generated transgenic mice expressing the ALS-associated FUS mutation, R521C, a point substitution in the middle of the FUS NLS. In the first of these, expression of mutant human, but not WT human FUS, caused the degeneration of motor axons, the cortex, and the hippocampus, resulting in progressive paralysis [130]. Interestingly, these ALS/FTLD phenotypes were also accompanied by ubiquitinated aggregates, though neither FUS nor TDP-43 were present within them. A study conducted 3 years later using the same R521C mutation found that an overly stable mutant FUS complex with brain-derived neurotrophic factor RNA was linked to neurological damage to dendrites and synapses [131]. Both studies managed to recapitulate a neurodegenerative phenotype exhibiting clear neuronal toxicity without any detectable cytoplasmic aggregation of TDP-43 or FUS, once more implicating RNA dysregulation as the true root of pathogenesis. Interestingly, an *in vitro* study utilizing patient fibroblasts expressing the exact same R521C mutation caused

extensive cytoplasmic FUS localization but also did not contain FUS inclusions under basal conditions [103]. Upon sodium arsenite treatment, however, the R521C fibroblasts exhibited FUS aggregation and colocalization with SGs whereas the WT did not [103]. Taken together, these data support the case that FUS mutations are capable of inducing neuronal toxicity without aggregation but that inclusions can still develop under external oxidative stress. This susceptibility will be discussed in more detail in Sect. 11.4.4.

One final note on FUS proteinopathies is that despite colocalization with common SG markers, pathological FUS granules may have unique properties. Yasuda and colleagues [132] recently reported that FUS facilitates preferential translation of RNA transcripts within granules that localize to cellular protrusions during cell migration. These RNA granules stained positive for both the SG marker TIA-1 and the tumor suppressor protein adenomatous polyposis coli (APC), but were shown to persist under cycloheximide treatment. Strikingly, tracking nascent protein synthesis with azidohomoalanine labeling revealed that APC-containing FUS granules were translationally active [132]. These data suggest that FUS granules may be distinct from endogenous SGs, which do not possess translation machinery and are sensitive to cycloheximide treatment. Notably, APC RNA granules could be induced by ALS associated FUS mutants, and tissue samples taken from FTLD-FUS patients immunostained positive for APC [132]. Taken together, these studies point towards a more complex picture for neurodegeneration wherein pathological inclusions may vary significantly in composition and behavior between disease subtypes. This variation warrants further characterization and poses a significant challenge to designing therapeutic interventions.

4.4 TDP-43, FUS, and Oxidative Stress

Besides disruptions to RNA homeostasis, a separate angle from which to consider the pathogenicity of TDP-43 and FUS (along with the role that SGs might play) stems from the unique susceptibility of neuronal cells to oxidative stress. First, the larger energy requirements of neurons might strain mitochondria, resulting in the production of more reactive oxygen species (ROS) and lowered mitochondrial efficiency, outcomes that are only exacerbated with age [133, 134]. Additionally, biomarkers for oxidative stress have long been identified in motor neurons and spinal cords of ALS patients [135]. These include a twofold increase in protein carbonyl levels [136], elevated levels of DNA 8-hydroxy-2′-deoxyguanosine (OH8dG) [137–139] and high immunoreactivity for lipid peroxidation and protein glycoxidation [140]. Mutations in other proteins implicated in pathogenesis might further predispose neurons to high levels of oxidative stress. For example, mutations in the chaperone protein VCP increase neuronal susceptibility to ROS build-up upon depletion of glutathione (GSH), a free radical scavenger which is also diminished in the motor cortices of ALS patients [135, 141, 142]. Interestingly, GSH depletion in NSC34 cells has been shown to result in the formation of insoluble inclusions containing the

hyperphosphorylated, truncated TDP-43 fragments commonly found in ALS histology [143]. Sensitivity to oxidative stress might also explain the age dependency of ALS, FTLD, and related neurological diseases. Given that SGs are a natural defense mechanism against sources of oxidative stress, the unresolved issue has been whether pathological inclusions constitute a root cause of pathogenesis, a symptom of RBP dysregulation, or a beneficial cellular response to exogenous stressors.

5 Concluding Remarks

Significant progress over the past decade has revealed that PrLDs/LCDs are critical for SG formation and that disruptions to these regions in ALS/FTLD patients are related to the development of pathological inclusions. To fulfill their functions as transient hubs of mRNP sorting and cellular signaling, SGs rely on dynamic, low affinity protein–protein and protein–RNA interactions. It is increasingly certain that these interactions are facilitated by PrLD-mediated oligomerization in conjunction with multivalent proteins, RNA, or other scaffolds mediated by post-translational modifications. SGs were first linked to neurological disease with the discovery that FUS and TDP-43, both RBPs with roles in mRNA metabolism, were frequent components of pathological inclusions found in ALS/FTLD patient histology. TDP-43 is now understood to be an endogenous SG component with a potential role in regulating mRNA triage and evidence that FUS also participates in cytoplasmic mRNP remodeling is growing. More importantly, both proteins robustly co-localize with SG markers in a wide range of cellular disease models, either due to cytoplasmic mislocalization, prolonged stress to which neurons are especially susceptible, or the mutation-induced enhancement of aggregation propensity. Whether this phenomenon plays a core role in neurodegenerative pathogenesis remains uncertain, largely due to the heterogeneous and highly variable nature of the disease class.

Interestingly, another SG RBP bearing prion-like properties has emerged as a candidate in neurodegenerative pathogenesis. Kim and colleagues [144, 145] showed that three distinct mutations in familial ALS pedigrees all resulted in a valine substitution in the center of the putative PrLD of the protein hnRNPA2B1, resulting in the similar pathological phenotype of mislocalization to the cytoplasm, aggregation into irreversible inclusions, and co-localization with SG markers. Interestingly, the valine mutations were computationally predicted to increase the prion-like propensity of the protein by enhancing homotypic oligomerization. Constructs of the hnRNPA2B1 PrLD formed amyloid fibers *in vitro*, and astoundingly, full length mutants fibrillized even quicker and were capable of hetero-oligomerization with other proteins [60, 145]. FUS-related proteins TAF15 and EWSR1 have also been recently linked to both ALS and FTLD [70, 146, 147]. Of note, these PrLD-mediated RNA aggregates are not restricted to ALS/FLTD pathologies, but have also been described in tauopathies, Alzheimer's and Parkinson's diseases [148–150].

Given the intrinsic link between pathological inclusions and SGs, one potential approach to therapeutic intervention might be the formulation of drugs to disrupt SG formation [65, 144, 151]. Because eIF2 alpha phosphorylation is known to induce SG assembly, a small molecule chemical which inhibits this pathway (GSK2606414) has been evaluated and shown to rescue TDP-43 toxicity in fly models and mammalian neurons [144]. Recent data suggest that aggregation due to oligomerization of PrLDs/LCDs is regulated by post-translational modifications within these protein domains [152–154]. By first identifying the enzymes responsible for these modifications, tissue-specific (in)activation of these proteins could provide new therapeutic avenues.

In addition, a disaggregase exhibiting remarkable effectiveness against misfolded protein aggregates, including those comprised of FUS, TDP-43, and TAF15, was engineered from the conserved yeast protein HSP104 [155]. Through screening of a random mutagenesis library of HSP104 variants and systematic optimization of the resulting potentiated variants, Shorter and colleagues abolished a putative autoinhibitory structural motif in the protein, reprogramming HSP104 with the capacity to solubilize TDP-43 and FUS aggregates, normalize proteotoxic mislocalization, and even rescue a neurological phenotype in nematode models [155]. An alternative strategy would center on how cytoplasmic aggregates are actively cleared. Genetic screens in yeast have revealed that different components in the autophagy pathway, such as VCP, are involved in such processes, consistent with recent genomic and biochemical data from ALS patients [156–158]. Indeed, VCP is known to play a role in returning TDP-43 back to solubility in the cytoplasm after stress release [159], and TDP-43 has even been shown to be preferentially cleared by autophagy [160–162]. Taken together, these data suggest a potential synergistic therapy by preventing prionogenic aggregation while enhancing autophagy. The ALS/FTLD disease system may display confounding heterogeneity, but where there are numerous openings to dysregulation there must also be diverse opportunities for intervention. Targeting prion-like interactions may constitute just one of many future complementary approaches.

Notes

Several key developments have advanced the SG field and strengthened the link between RBP dysregulation and ALS/FTLD pathology since the acceptance of this manuscript. Noting that SG-associated mRNPs remain stable ex vivo, Jain et al. have isolated these "SG cores" for proteomic analyses, finding that ATPases are critical for SG formation and disassembly [163]. They propose that these cores are enveloped by a more dynamic "shell" governed by protein–protein interactions between IDRs (see Sect. 3.2). Additionally, emergent data indicate that RNA accelerates IDR-mediated phase separation [164] and that specific sequences can influence the biophysical properties of mRNP droplets, suggesting that mRNA may encode "architectural determinants for various non-membranous organelles" ([165];

reviewed recently by [166]). Work also continues on pinpointing mechanisms of toxicity in the ALS/FTLD disease spectrum. ALS-associated hnRNPA1 and FUS mutants fibrillize more readily than WT ex vivo and mutant FUS expression in Xenopus retinal neurons attenuates protein synthesis [167, 168]. Taken together, these data reaffirm the hypothesis that demixing events mediated by IDRs, which occur naturally in the cellular stress response, may actually seed pathological aggregates. How these aggregates disrupt RNA metabolism is a question of growing importance.

Acknowledgements We would like to thank Drs. Phillip Sharp, Nancy Kedersha, Anaïs Aulas, and Voula Mili for critical reading of the manuscript, Drs. Steve McKnight and Masato Kato on sharing their insights on the role of low-complexity domains in RNA granule formation. This work was partly supported by Johns Hopkins Catalyst Award and NIH R01-GM104135 to A.K.L.L.

References

1. Höck J, Weinmann L, Ender C et al (2007) Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep 8:1052–1060. doi:10.1038/sj.embor.7401088
2. La Rocca G, Olejniczak SH, González AJ et al (2015) In vivo, Argonaute-bound microRNAs exist predominantly in a reservoir of low molecular weight complexes not associated with mRNA. Proc Natl Acad Sci U S A 112:767–772. doi:10.1073/pnas.1424217112
3. Bono F, Gehring NH (2011) Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA Biol 8:24–29
4. Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. Nat Rev Genet 13:246–259. doi:10.1038/nrg3160
5. Anderson P, Kedersha NL (2009) RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat Rev Mol Cell Biol 10:430–436. doi:10.1038/nrm2694
6. Anderson P, Kedersha NL (2006) RNA granules. J Cell Biol 172:803–808. doi:10.1083/jcb.200512082
7. Buchan JR, Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. Mol Cell 36:932–941. doi:10.1016/j.molcel.2009.11.020
8. Spector DL (2006) SnapShot: cellular bodies. Cell 127:1071. doi:10.1016/j.cell.2006.11.026
9. Arrigo AP, Suhan JP, Welch WJ (1988) Dynamic changes in the structure and intracellular locale of the mammalian low-molecular-weight heat shock protein. Mol Cell Biol 8:5059–5071
10. Collier NC, Heuser J, Levy MA, Schlesinger MJ (1988) Ultrastructural and biochemical analysis of the stress granule in chicken embryo fibroblasts. J Cell Biol 106:1131–1139
11. Collier NC, Schlesinger MJ (1986) The dynamic state of heat shock proteins in chicken embryo fibroblasts. J Cell Biol 103:1495–1507
12. Kedersha NL, Gupta M, Li W et al (1999) RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. J Cell Biol 147:1431–1442
13. Buchan JR, Muhlrad D, Parker R (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. J Cell Biol 183:441–455. doi:10.1083/jcb.200807043
14. Farny NG, Kedersha NL, Silver PA (2009) Metazoan stress granule assembly is mediated by P-eIF2alpha-dependent and -independent mechanisms. RNA 15:1814–1821. doi:10.1261/rna.1684009

15. Grousl T, Ivanov P, Frydlova I et al (2009) Robust heat shock induces eIF2 -phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. J Cell Sci 122:2078–2088. doi:10.1242/jcs.045104
16. Hoyle NP, Castelli LM, Campbell SG et al (2007) Stress-dependent relocalization of translationally primed mRNPs to cytoplasmic granules that are kinetically and spatially distinct from P-bodies. J Cell Biol 179:65–74. doi:10.1083/jcb.200707010
17. Souquere S, Mollet S, Kress M et al (2009) Unravelling the ultrastructure of stress granules and associated P-bodies in human cells. J Cell Sci 122:3619–3626. doi:10.1242/jcs.054437
18. Anderson P, Kedersha NL (2008) Stress granules: the Tao of RNA triage. Trends Biochem Sci 33:141–150. doi:10.1016/j.tibs.2007.12.003
19. Kedersha NL, Anderson P (2007) Mammalian stress granules and processing bodies. Methods Enzymol 431:61–81. doi:10.1016/S0076-6879(07)31005-7
20. Moeller BJ, Cao Y, Li CY, Dewhirst MW (2004) Radiation activates HIF-1 to regulate vascular radiosensitivity in tumors: role of reoxygenation, free radicals, and stress granules. Cancer Cell 5:429–441
21. White JP, Lloyd RE (2012) Regulation of stress granules in virus systems. Trends Microbiol 20:175–183. doi:10.1016/j.tim.2012.02.001
22. Warner JR, Rich A, Hall CE (1962) Electron microscope studies of ribosomal clusters synthesizing hemoglobin. Science 138:1399–1403. doi:10.1126/science.138.3548.1399
23. Mokas S, Mills JR, Garreau C et al (2009) Uncoupling stress granule assembly and translation initiation inhibition. Mol Biol Cell 20:2673–2683. doi:10.1091/mbc.E08-10-1061
24. Baguet A, Degot S, Cougot N et al (2007) The exon-junction-complex-component metastatic lymph node 51 functions in stress-granule assembly. J Cell Sci 120:2774–2784. doi:10.1242/jcs.009225
25. Kedersha NL, Stoecklin G, Ayodele M et al (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. J Cell Biol 169:871–884. doi:10.1083/jcb.200502088
26. Mazroui R, Huot M-E, Tremblay S et al (2002) Trapping of messenger RNA by Fragile X Mental Retardation protein into cytoplasmic granules induces translation repression. Hum Mol Genet 11:3007–3017
27. Wilczynska A, Aigueperse C, Kress M et al (2005) The translational regulator CPEB1 provides a link between dcp1 bodies and stress granules. J Cell Sci 118:981–992. doi:10.1242/jcs.01692
28. Kedersha NL, Ivanov P, Anderson P (2013) Stress granules and cell signaling: more than just a passing phase? Trends Biochem Sci. doi:10.1016/j.tibs.2013.07.004
29. Piotrowska J, Hansen SJ, Park N et al (2010) Stable formation of compositionally unique stress granules in virus-infected cells. J Virol 84:3654–3665. doi:10.1128/JVI.01320-09
30. Buchan JR, Yoon J-H, Parker R (2011) Stress-specific composition, assembly and kinetics of stress granules in saccharomyces cerevisiae. J Cell Sci 124:228–239. doi:10.1242/jcs.078444
31. Stoecklin G, Stubbs T, Kedersha NL et al (2004) MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay. EMBO J 23:1313–1324. doi:10.1038/sj.emboj.7600163
32. Gilks N, Kedersha NL, Ayodele M et al (2004) Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol Biol Cell 15:5383–5398. doi:10.1091/mbc.E04-08-0715
33. Tanaka T, Ohashi S, Kobayashi S (2014) Roles of YB-1 under arsenite-induced stress: translational activation of HSP70 mRNA and control of the number of stress granules. Biochim Biophys Acta 1840:985–992. doi:10.1016/j.bbagen.2013.11.002
34. Mollet S, Cougot N, Wilczynska A et al (2008) Translationally repressed mRNA transiently cycles through stress granules during stress. Mol Biol Cell 19:4469–4479. doi:10.1091/mbc.E08-05-0499
35. Zhang J, Okabe K, Tani T, Funatsu T (2011) Dynamic association-dissociation and harboring of endogenous mRNAs in stress granules. J Cell Sci 124:4087–4095. doi:10.1242/jcs.090951

36. Zhang T, Mullane PC, Periz G, Wang J (2011) TDP-43 neurotoxicity and protein aggregation modulated by heat shock factor and insulin/IGF-1 signaling. Hum Mol Genet 20:1952–1965
37. Kedersha NL, Anderson P (2002) Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. Biochem Soc Trans 30:963–969
38. Kedersha NL, Cho MR, Li W et al (2000) Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. J Cell Biol 151:1257–1268
39. Hilliker A, Gao Z, Jankowsky E, Parker R (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. Mol Cell 43:962–972. doi:10.1016/j.molcel.2011.08.008
40. Shih J-W, Tsai T-Y, Chao C-H, Wu Lee Y-H (2008) Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein. Oncogene 27:700–714. doi:10.1038/sj.onc.1210687
41. Arimoto K, Fukuda H, Imajoh-Ohmi S et al (2008) Formation of stress granules inhibits apoptosis by suppressing stress-responsive MAPK pathways. Nat Cell Biol 10:1324–1332. doi:10.1038/ncb1791
42. Eisinger-Mathason TSK, Andrade J, Groehler AL et al (2008) Codependent functions of RSK2 and the apoptosis-promoting factor TIA-1 in stress granule assembly and cell survival. Mol Cell 31:722–736. doi:10.1016/j.molcel.2008.06.025
43. Gareau C, Fournier M-J, Filion C et al (2011) p21(WAF1/CIP1) upregulation through the stress granule-associated protein CUGBP1 confers resistance to bortezomib-mediated apoptosis. PLoS One 6, e20254. doi:10.1371/journal.pone.0020254
44. Thedieck K, Holzwarth B, Prentzell MT et al (2013) Inhibition of mTORC1 by astrin and stress granules prevents apoptosis in cancer cells. Cell 154:859–874. doi:10.1016/j.cell.2013.07.031
45. Kim WJ, Back SH, Kim V et al (2005) Sequestration of TRAF2 into stress granules interrupts tumor necrosis factor signaling under stress conditions. Mol Cell Biol 25:2450–2462. doi:10.1128/MCB.25.6.2450-2462.2005
46. Tsai N-P, Wei L-N (2010) RhoA/ROCK1 signaling regulates stress granule formation and apoptosis. Cell Signal 22:668–675. doi:10.1016/j.cellsig.2009.12.001
47. Li W, Simarro M, Kedersha NL, Anderson P (2004) FAST is a survival protein that senses mitochondrial stress and modulates TIA-1-regulated changes in protein expression. Mol Cell Biol 24:10718–10732. doi:10.1128/MCB.24.24.10718-10732.2004
48. Kwon S, Zhang Y, Matthias P (2007) The deacetylase HDAC6 is a novel critical component of stress granules involved in the stress response. Genes Dev 21:3381–3394. doi:10.1101/gad.461107
49. Fournier M-J, Coudert L, Mellaoui S et al (2013) Inactivation of the mTORC1-eukaryotic translation initiation factor 4E pathway alters stress granule formation. Mol Cell Biol 33:2285–2301. doi:10.1128/MCB.01517-12
50. Hofmann S, Cherkasova V, Bankhead P et al (2012) Translation suppression promotes stress granule formation and cell survival in response to cold shock. Mol Biol Cell 23:3786–3800. doi:10.1091/mbc.E12-04-0296
51. Wippich F, Bodenmiller B, Trajkovska MG et al (2013) Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. Cell 152:791–805. doi:10.1016/j.cell.2013.01.033
52. Fournier M-J, Gareau C, Mazroui R (2010) The chemotherapeutic agent bortezomib induces the formation of stress granules. Cancer Cell Int 10:12. doi:10.1186/1475-2867-10-12
53. Li YR, King OD, Shorter J, Gitler AD (2013) Stress granules as crucibles of ALS pathogenesis. J Cell Biol 201:361–372. doi:10.1083/jcb.201302044
54. Onomoto K, Yoneyama M, Fung G et al (2014) Antiviral innate immunity and stress granule responses. Trends Immunol. doi:10.1016/j.it.2014.07.006
55. Banjade S, Rosen MK (2014) Phase transitions of multivalent proteins can promote clustering of membrane receptors. Elife. doi:10.7554/eLife.04123
56. Li P, Banjade S, Cheng H-C et al (2012) Phase transitions in the assembly of multivalent signalling proteins. Nature 483:336–340. doi:10.1038/nature10879

57. Brangwynne CP (2013) Phase transitions and size scaling of membrane-less organelles. J Cell Biol 203:875–881. doi:10.1083/jcb.201308087
58. Hyman AA, Simons K (2012) Cell biology. Beyond oil and water—phase transitions in cells. Science 337:1047–1049. doi:10.1126/science.1223728
59. Weber SC, Brangwynne CP (2012) Getting RNA and protein in phase. Cell 149:1188–1191. doi:10.1016/j.cell.2012.05.022
60. Kato M, Han TW, Xie S et al (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell 149:753–767. doi:10.1016/j.cell.2012.04.017
61. Schwartz JC, Wang X, Podell ER, Cech TR (2013) RNA seeds higher-order assembly of FUS protein. Cell Rep 5:918–925. doi:10.1016/j.celrep.2013.11.017
62. Hyman AA, Weber CA, Jülicher F (2014) Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol 30:39–58. doi:10.1146/annurev-cellbio-100913-013325
63. Shevtsov SP, Dundr M (2011) Nucleation of nuclear bodies by RNA. Nat Cell Biol 13:167–173. doi:10.1038/ncb2157
64. Teixeira D, Sheth U, Valencia-Sanchez MA et al (2005) Processing bodies require RNA for assembly and contain nontranslating mRNAs. RNA 11:371–382. doi:10.1261/rna.7258505
65. Leung AKL (2014) Poly(ADP-ribose): an organizer of cellular architecture. J Cell Biol 205:613–619. doi:10.1083/jcb.201402114
66. Leung AKL, Vyas S, Rood JE et al (2011) Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. Mol Cell 42:489–499. doi:10.1016/j.molcel.2011.04.015
67. Shorter J, Lindquist S (2005) Prions as adaptive conduits of memory and inheritance. Nat Rev Genet 6:435–450. doi:10.1038/nrg1616
68. King OD, Gitler AD, Shorter J (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res 1462:61–80. doi:10.1016/j.brainres.2012.01.016
69. Alberti S, Halfmann R, King O et al (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell 137:146–158. doi:10.1016/j.cell.2009.02.044
70. Couthouis J, Hart MP, Shorter J et al (2011) A yeast functional screen predicts new candidate ALS disease genes. Proc Natl Acad Sci U S A 108:20881–20890. doi:10.1073/pnas.1109434108
71. Han TW, Kato M, Xie S et al (2012) Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell 149:768–779. doi:10.1016/j.cell.2012.04.016
72. Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A 91:5355–5358
73. Tourrière H, Chebli K, Zekri L et al (2003) The RasGAP-associated endoribonuclease G3BP assembles stress granules. J Cell Biol 160:823–831. doi:10.1083/jcb.200212128
74. Glover JR, Lindquist S (1998) Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell 94:73–82
75. Li X, Rayman JB, Kandel ER, Derkatch IL (2014) Functional role of Tia1/Pub1 and Sup35 prion domains: directing protein synthesis machinery to the tubulin cytoskeleton. Mol Cell. doi:10.1016/j.molcel.2014.05.027
76. Boyault C, Zhang Y, Fritah S et al (2007) HDAC6 controls major cell response pathways to cytotoxic accumulation of protein aggregates. Genes Dev 21:2172–2181. doi:10.1101/gad.436407
77. Rikhvanov EG, Romanova NV, Chernoff YO (2007) Chaperone effects on prion and non-prion aggregates. Prion 1:217–222
78. Guil S, Long JC, Cáceres JF (2006) hnRNP A1 relocalization to the stress granules reflects a role in the stress response. Mol Cell Biol 26:5744–5758. doi:10.1128/MCB.00224-06

79. Schmidlin M, Lu M, Leuenberger SA et al (2004) The ARE-dependent mRNA-destabilizing activity of BRF1 is regulated by protein kinase B. EMBO J 23:4760–4769. doi:10.1038/sj.emboj.7600477
80. Yoon J-H, Abdelmohsen K, Srikantan S et al (2013) Tyrosine phosphorylation of HuR by JAK3 triggers dissociation and degradation of HuR target mRNAs. Nucleic Acids Res. doi:10.1093/nar/gkt903
81. Hottiger MO, Hassa PO, Lüscher B et al (2010) Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends Biochem Sci. doi:10.1016/j.tibs.2009.12.003
82. van der Lee R, Buljan M, Lang B et al (2014) Classification of intrinsically disordered regions and proteins. Chem Rev 114:6589–6631. doi:10.1021/cr400525m
83. Rowland LP, Shneider NA (2001) Amyotrophic lateral sclerosis. N Engl J Med 344:1688–1700. doi:10.1056/NEJM200105313442207
84. Rabinovici GD, Miller BL (2010) Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. CNS Drugs 24:375–398. doi:10.2165/11533100-000000000-00000
85. Ling S-C, Polymenidou M, Cleveland DW (2013) Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron 79:416–438. doi:10.1016/j.neuron.2013.07.033
86. Robberecht W, Philips T (2013) The changing scene of amyotrophic lateral sclerosis. Nat Rev Neurosci 14:248–264. doi:10.1038/nrn3430
87. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314(5796):130–133
88. Tollervey JR, Curk T, Rogelj B et al (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci 14:452–458. doi:10.1038/nn.2778
89. Lagier-Tourenne C, Polymenidou M, Hutt KR et al (2012) Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat Neurosci 15:1488–1497. doi:10.1038/nn.3230
90. McDonald KK, Aulas A, Destroismaisons L et al (2011) TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum Mol Genet 20:1400–1410. doi:10.1093/hmg/ddr021
91. Liu-Yesucevitz L, Bilgutay A, Zhang Y-J et al (2010) Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PLoS One 5, e13250. doi:10.1371/journal.pone.0013250
92. Dewey CM, Cenik B, Sephton CF et al (2011) TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. Mol Cell Biol 31:1098–1108. doi:10.1128/MCB.01279-10
93. Bentmann E, Neumann M, Tahirovic S et al (2012) Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). J Biol Chem 287:23079–23094. doi:10.1074/jbc.M111.328757
94. Liu-Yesucevitz L, Lin AY, Ebata A et al (2014) ALS-linked mutations enlarge TDP-43-enriched neuronal RNA granules in the dendritic arbor. J Neurosci 34:4167–4174. doi:10.1523/JNEUROSCI.2350-13.2014
95. Chiò A, Calvo A, Mazzini L et al (2012) Extensive genetics of ALS: a population-based study in Italy. Neurology 79:1983–1989. doi:10.1212/WNL.0b013e3182735d36
96. Aulas A, Stabile S, Vande Velde C (2012) Endogenous TDP-43, but not FUS, contributes to stress granule assembly via G3BP. Mol Neurodegener 7:54. doi:10.1186/1750-1326-7-54
97. Aulas A, Caron G, Gkogkas CG et al (2015) G3BP1 promotes stress-induced RNA granule interactions to preserve polyadenylated mRNA. J Cell Biol 209:73–84. doi:10.1083/jcb.201408092
98. Iko Y, Kodama TS, Kasai N et al (2004) Domain architectures and characterization of an RNA-binding protein, TLS. J Biol Chem 279:44834–44840. doi:10.1074/jbc.M408552200

99. Dormann D, Rodde R, Edbauer D et al (2010) ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. EMBO J 29:2841–2857. doi:10.1038/emboj.2010.143
100. Gal J, Zhang J, Kwinter DM et al (2011) Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. Neurobiol Aging 32:2323.e27–2323.e40. doi:10.1016/j.neurobiolaging.2010.06.010
101. Bosco DA, Lemay N, Ko HK et al (2010) Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. Hum Mol Genet 19:4160–4175. doi:10.1093/hmg/ddq335
102. Ito D, Suzuki N (2011) Conjoint pathologic cascades mediated by ALS/FTLD-U linked RNA-binding proteins TDP-43 and FUS. Neurology 77(17):1636–1643
103. Vance C, Scotter EL, Nishimura AL et al (2013) ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. Hum Mol Genet 22:2676–2688. doi:10.1093/hmg/ddt117
104. Daigle JG, Lanson NA, Smith RB et al (2013) RNA-binding ability of FUS regulates neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations. Hum Mol Genet 22:1193–1205. doi:10.1093/hmg/dds526
105. Kino Y, Washizu C, Aquilanti E et al (2011) Intracellular localization and splicing regulation of FUS/TLS are variably affected by amyotrophic lateral sclerosis-linked mutations. Nucleic Acids Res 39:2781–2798. doi:10.1093/nar/gkq1162
106. Andersson MK, Ståhlberg A, Arvidsson Y et al (2008) The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. BMC Cell Biol 9:37. doi:10.1186/1471-2121-9-37
107. Blechingberg J, Luo Y, Bolund L et al (2012) Gene expression responses to FUS, EWS, and TAF15 reduction and stress granule sequestration analyses identifies FET-protein non-redundant functions. PLoS One 7, e46251. doi:10.1371/journal.pone.0046251
108. Sama RRK, Ward CL, Kaushansky LJ et al (2013) FUS/TLS assembles into stress granules and is a prosurvival factor during hyperosmolar stress. J Cell Physiol 228:2222–2231. doi:10.1002/jcp.24395
109. Zhang ZC, Chook YM (2012) Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). Proc Natl Acad Sci U S A 109:12017–12021. doi:10.1073/pnas.1207247109
110. Dormann D, Madl T, Valori CF et al (2012) Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. EMBO J 31:4258–4275. doi:10.1038/emboj.2012.261
111. Johnson BS, Snead D, Lee JJ et al (2009) TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. J Biol Chem 284:20329–20339. doi:10.1074/jbc.M109.010264
112. Cushman M, Johnson BS, King OD et al (2010) Prion-like disorders: blurring the divide between transmissibility and infectivity. J Cell Sci 123:1191–1201. doi:10.1242/jcs.051672
113. Fushimi K, Long C, Jayaram N et al (2011) Expression of human FUS/TLS in yeast leads to protein aggregation and cytotoxicity, recapitulating key features of FUS proteinopathy. Protein Cell 2:141–149. doi:10.1007/s13238-011-1014-5
114. Sun Z, Diaz Z, Fang X et al (2011) Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS Biol 9, e1000614. doi:10.1371/journal.pbio.1000614
115. Barmada SJ, Skibinski G, Korb E et al (2010) Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. J Neurosci 30:639–649. doi:10.1523/JNEUROSCI.4988-09.2010
116. Ling S-C, Albuquerque CP, Han JS et al (2010) ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc Natl Acad Sci U S A 107:13318–13323. doi:10.1073/pnas.1008227107

117. Johnson BS, McCaffery JM, Lindquist S, Gitler AD (2008) A yeast TDP-43 proteinopathy model: exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. Proc Natl Acad Sci U S A 105:6439–6444. doi:10.1073/pnas.0802082105
118. Kabashi E, Lin L, Tradewell ML et al (2010) Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. Hum Mol Genet 19:671–683. doi:10.1093/hmg/ddp534
119. Lanson NA, Maltare A, King H et al (2011) A Drosophila model of FUS-related neurodegeneration reveals genetic interaction between FUS and TDP-43. Hum Mol Genet 20:2510–2523. doi:10.1093/hmg/ddr150
120. Polymenidou M, Cleveland DW (2011) The seeds of neurodegeneration: prion-like spreading in ALS. Cell 147:498–508. doi:10.1016/j.cell.2011.10.011
121. Meyerowitz J, Parker SJ, Vella LJ et al (2011) C-Jun N-terminal kinase controls TDP-43 accumulation in stress granules induced by oxidative stress. Mol Neurodegener 6:57. doi:10.1186/1750-1326-6-57
122. Parker SJ, Meyerowitz J, James JL et al (2012) Endogenous TDP-43 localized to stress granules can subsequently form protein aggregates. Neurochem Int 60:415–424. doi:10.1016/j.neuint.2012.01.019
123. Kim SH, Shanware NP, Bowler MJ, Tibbetts RS (2010) Amyotrophic lateral sclerosis-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to co-regulate HDAC6 mRNA. J Biol Chem 285:34097–34105. doi:10.1074/jbc.M110.154831
124. Polymenidou M, Lagier-Tourenne C, Hutt KR et al (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci 14:459–468. doi:10.1038/nn.2779
125. Arnold ES, Ling S-C, Huelga SC et al (2013) ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. Proc Natl Acad Sci U S A 110:E736–E745. doi:10.1073/pnas.1222809110
126. Igaz LM, Kwong LK, Lee EB et al (2011) Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. J Clin Invest 121:726–738. doi:10.1172/JCI44867
127. Hicks GG, Singh N, Nashabi A et al (2000) Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. Nat Genet 24:175–179. doi:10.1038/72842
128. Sasayama H, Shimamura M, Tokuda T et al (2012) Knockdown of the Drosophila fused in sarcoma (FUS) homologue causes deficient locomotive behavior and shortening of motoneuron terminal branches. PLoS One 7, e39483. doi:10.1371/journal.pone.0039483
129. Chen Y, Yang M, Deng J et al (2011) Expression of human FUS protein in Drosophila leads to progressive neurodegeneration. Protein Cell 2:477–486. doi:10.1007/s13238-011-1065-7
130. Huang C, Zhou H, Tong J et al (2011) FUS transgenic rats develop the phenotypes of amyotrophic lateral sclerosis and frontotemporal lobar degeneration. PLoS Genet 7, e1002011. doi:10.1371/journal.pgen.1002011
131. Qiu H, Lee S, Shang Y et al (2014) ALS-associated mutation FUS-R521C causes DNA damage and RNA splicing defects. J Clin Invest 124:981–999. doi:10.1172/JCI72723
132. Yasuda K, Zhang H, Loiselle D et al (2013) The RNA-binding protein Fus directs translation of localized mRNAs in APC-RNP granules. J Cell Biol 203:737–746. doi:10.1083/jcb.201306058
133. Barber SC, Shaw PJ (2010) Oxidative stress in ALS: key role in motor neuron injury and therapeutic target. Free Radic Biol Med 48:629–641. doi:10.1016/j.freeradbiomed.2009.11.018
134. Beal MF (2002) Oxidatively modified proteins in aging and disease. Free Radic Biol Med 32:797–803
135. Carrì MT, Valle C, Bozzo F, Cozzolino M (2015) Oxidative stress and mitochondrial damage: importance in non-SOD1 ALS. Front Cell Neurosci 9:41. doi:10.3389/fncel.2015.00041
136. Shaw PJ, Ince PG, Falkous G, Mantle D (1995) Oxidative damage to protein in sporadic motor neuron disease spinal cord. Ann Neurol 38:691–695. doi:10.1002/ana.410380424
137. Bogdanov M, Brown RH, Matson W et al (2000) Increased oxidative damage to DNA in ALS patients. Free Radic Biol Med 29:652–658

138. Ferrante RJ, Browne SE, Shinobu LA et al (1997) Evidence of increased oxidative damage in both sporadic and familial amyotrophic lateral sclerosis. J Neurochem 69:2064–2074
139. Ihara Y, Nobukuni K, Takata H, Hayabara T (2005) Oxidative stress and metal content in blood and cerebrospinal fluid of amyotrophic lateral sclerosis patients with and without a Cu, Zn-superoxide dismutase mutation. Neurol Res 27:105–108. doi:10.1179/016164105X18430
140. Shibata N, Nagai R, Uchida K et al (2001) Morphological evidence for lipid peroxidation and protein glycoxidation in spinal cords from sporadic amyotrophic lateral sclerosis patients. Brain Res 917:97–104
141. Hirano M et al. (2015) VCP gene analyses in Japanese patients with sporadic amyotrophic lateral sclerosis identify a new mutation. Neurobiol. Aging 36: 1604.e1–6
142. Weiduschat N et al (2014) Motor cortex glutathione deficit in ALS measured in vivo with the J-editing technique. Neurosci Lett 570:102–107
143. Iguchi Y, Katsuno M, Takagi S et al (2012) Oxidative stress induced by glutathione depletion reproduces pathological modifications of TDP-43 linked to TDP-43 proteinopathies. Neurobiol Dis 45:862–870. doi:10.1016/j.nbd.2011.12.002
144. Kim H-J, Raphael AR, LaDow ES et al (2013) Therapeutic modulation of eIF2α phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. Nat Genet. doi:10.1038/ng.2853
145. Kim HJ, Kim NC, Wang Y-D et al (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 495:467–473. doi:10.1038/nature11922
146. Couthouis J, Hart MP, Erion R et al (2012) Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum Mol Genet 21:2899–2911. doi:10.1093/hmg/dds116
147. Mackenzie IRA, Neumann M (2012) FET proteins in frontotemporal dementia and amyotrophic lateral sclerosis. Brain Res 1462:40–43. doi:10.1016/j.brainres.2011.12.010
148. Ash PEA, Vanderweyde TE, Youmans KL et al (2014) Pathological stress granules in Alzheimer's disease. Brain Res 1584:52–58. doi:10.1016/j.brainres.2014.05.052
149. Banks GT, Kuta A, Isaacs AM, Fisher EMC (2008) TDP-43 is a culprit in human neurodegeneration, and not just an innocent bystander. Mamm Genome 19:299–305. doi:10.1007/s00335-008-9117-x
150. Vanderweyde T, Yu H, Varnum M et al (2012) Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. J Neurosci 32:8270–8283. doi:10.1523/JNEUROSCI.1592-12.2012
151. Sidrauski C, McGeachy AM, Ingolia NT, Walter P (2015) The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly. Elife. doi:10.7554/eLife.05033
152. Gambetta MC, Müller J (2014) O-GlcNAcylation prevents aggregation of the Polycomb group repressor Polyhomeotic. Dev Cell 31:629–639. doi:10.1016/j.devcel.2014.10.020
153. Kwon I, Kato M, Xiang S et al (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell 155:1049–1060. doi:10.1016/j.cell.2013.10.033
154. Kwon I, Xiang S, Kato M et al (2014) Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science. doi:10.1126/science.1254917
155. Jackrel ME, DeSantis ME, Martinez BA et al (2014) Potentiated Hsp104 variants antagonize diverse proteotoxic misfolding events. Cell 156:170–182. doi:10.1016/j.cell.2013.11.047
156. Buchan JR, Kolaitis R-M, Taylor JP, Parker R (2013) Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. Cell 153:1461–1474. doi:10.1016/j.cell.2013.05.037
157. Cirulli ET, Lasseigne BN, Petrovski S et al (2015) Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science. doi:10.1126/science.aaa3650
158. Meyer H, Weihl CC (2014) The VCP/p97 system at a glance: connecting cellular function to disease pathogenesis. J Cell Sci 127:3877–3883. doi:10.1242/jcs.093831
159. Barmada SJ, Serio A, Arjun A et al (2014) Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. Nat Chem Biol 10:677–685. doi:10.1038/nchembio.1563

160. Wang I-F, Guo B-S, Liu Y-C et al (2012) Autophagy activators rescue and alleviate pathogenesis of a mouse model with proteinopathies of the TAR DNA-binding protein 43. Proc Natl Acad Sci U S A 109:15024–15029. doi:10.1073/pnas.1206362109
161. Wang I-F, Tsai K-J, Shen C-KJ (2013) Autophagy activation ameliorates neuronal pathogenesis of FTLD-U mice: a new light for treatment of TARDBP/TDP-43 proteinopathies. Autophagy 9:239–240. doi:10.4161/auto.22526
162. Wang X, Fan H, Ying Z et al (2010) Degradation of TDP-43 and its pathogenic form by autophagy and the ubiquitin-proteasome system. Neurosci Lett 469:112–116. doi:10.1016/j.neulet.2009.11.055
163. Jain S, Wheeler JR, Walters RW et al. (2016) ATPase-modulated stress granules contain a diverse proteome and substructure. Cell 164:487–498. doi:10.1016/j.cell.2015.12.038.
164. Lin Y, Protter DSW, Rosen MK, Parker R (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. Mol Cell 60:208–219. doi:10.1016/j.molcel.2015.08.018.
165. Zhang H, Elbaum-Garfinkle S, Langdon EM et al. (2015) RNA controls PolyQ protein phase transitions. Mol Cell 60:220–230. doi:10.1016/j.molcel.2015.09.017.
166. Guo L, Shorter J (2015) It's raining liquids: RNA tunes viscoelasticity and dynamics of membraneless organelles. Mol Cell 60:189–192. doi:10.1016/j.molcel.2015.10.006.
167. Molliex A, Temirov J, Lee J et al. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell 163:123–133. doi:10.1016/j.cell.2015.09.015.
168. Murakami T, Qamar S, Lin JQ et al. (2015) ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. Neuron 88:678–690. doi:10.1016/j.neuron.2015.10.030.

Chapter 12
Post-Translational Modifications and RNA-Binding Proteins

Michael T. Lovci, Mario H. Bengtson, and Katlin B. Massirer

Abstract RNA-binding proteins affect cellular metabolic programs through development and in response to cellular stimuli. Though much work has been done to elucidate the roles of a handful of RNA-binding proteins and their effect on RNA metabolism, the progress of studies to understand the effects of post-translational modifications of this class of proteins is far from complete. This chapter summarizes the work that has been done to identify the consequence of post-translational modifications to some RNA-binding proteins. The effects of these modifications have been shown to increase the panoply of functions that a given RNA-binding protein can assume. We will survey the experimental methods that are used to identify the presence of several protein modifications and methods that attempt to discern the consequence of these modifications.

Keywords RNA-binding proteins • Post-translational modifications • SUMOylation • Ubiquitination • Phosphorylation

1 Introduction

RNA-binding proteins (RBPs) regulate RNAs at every stage of their existence. This includes processes that govern RNA metabolism from capping and polynucleotide extension, RNA splicing, subcellular RNA localization, cellular export, translation (initiation elongation and extension), to RNA destruction. This class of ~1200–1600 proteins has important roles in the etiology of disease and therefore advances in the understanding of these proteins hold the promise to be directly

M.T. Lovci • K.B. Massirer (✉)
Center for Molecular Biology and Genetic Engineering,
University of Campinas, CBMEG-UNICAMP, Av Candido Rondon 400,
Campinas, Sao Paulo 13083-875, Brazil
e-mail: kmassire@unicamp.br

M.H. Bengtson
Department of Biochemistry and Tissue Biology, Institute of Biology,
University of Campinas, UNICAMP, Av Monteiro Lobato 255,
Campinas, Sao Paulo 13083-970, Brazil

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7_12

applicable to the treatment of human neurodegenerative diseases, cancers and developmental disorders [1–5]. It has been known for some time that mRNA splicing is coupled to signal transduction and posttranslational modifications (PTMs) [6]. A full understanding of this intertwined network of processes has been complicated by the realization that RNA-binding proteins are a diverse class of regulators which themselves undergo extensive regulation via splicing, alternative 5′ and 3′ ends and various post-translational modifications.

Post-translational modifications follow from various signaling pathways to cause activation of enzymes that add or remove PTM moieties (Fig. 12.1a). The set of post-translational modifications known to affect RNA-binding protein function includes at least: the reversible addition/removal of phosphate groups (PO_3) by kinases/phosphatases, of methyl groups (CH_3) by methylases/demethylases, of acetyl groups (C_2H_30) by acetylases/deacetylases, of the small protein ubiquitin (~8.5 kDa protein) by ubiquitin ligases/deubiquitinating enzymes, of SUMO

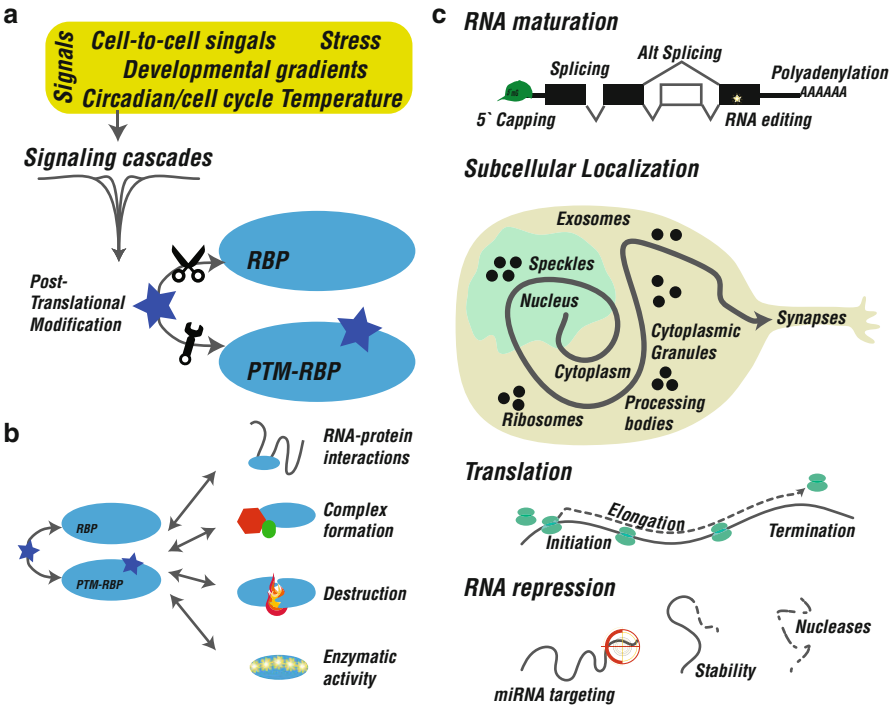


Fig. 12.1 Schematic of the effects of PTM on RNA-binding proteins. (**a**) Signal integration. Various signals from inside a cell and from external sources activate signaling cascades that converge on the regulators of PTM state. (**b**) PTM may activate or deactivate the functions of an RNA-binding protein, including altering RNA targets, protein partners, mediating protein degradation or intrinsic enzymatic activities. (**c**) The altered functions of RNA-binding proteins lead to overall differences in the metabolism of RNAs at every stage of their existence, from transcription through destruction

(~12 kDa proteins) by ubiquitin ligases/SUMO proteases, and of glycans (polysaccharides) by glycosyltransferases/exoglycosidases or proteolytic cleavage by proteases. Although classically ubiquitination is associated with proteassomal degradation, some studies point to other functional roles for ubiquitin conjugation including localization and regulating protein interaction partners. Further, there have been observations of functional differences in the activity of polyubiquitin chains, depending on which lysine position links the ubiquitin monomers [7].

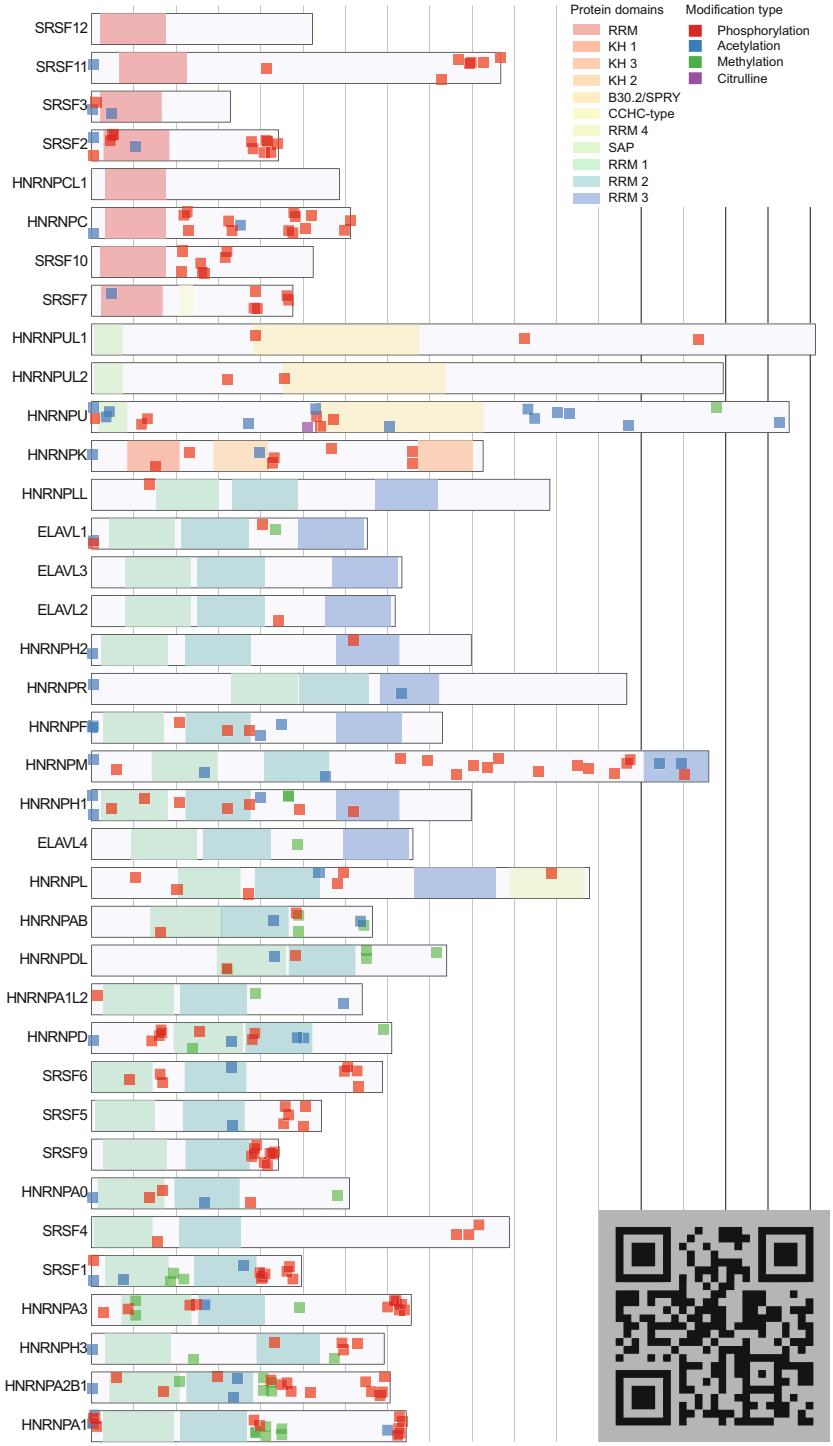
RBPs are affected by these PTMs in diverse and currently unpredictable ways. RBP subcellular localization, affinity for RNA, enzymatic activity and association with cofactors have all been shown to be directed by PTM state (Fig. 12.1b). Since these consequences on protein function on RNA fate are often indistinguishable without close inspection, the exact stage where a PTM effects change is sometimes unclear. We have organized this chapter following the life-cycle of an mRNA from transcription through splicing, then nuclear export, subcellular targeting, translation and ultimately destruction. We highlight research at each stage of the RNA life-cycle that shows the indispensability of PTMs for fidelity of RNA regulation (Fig. 12.1c). Where possible we point to studies that discuss the exact effect of PTMs and describe experiments that can reveal this information. Overall, this chapter aims to be a review of the work done to bridge the gap between proteomics and transcriptomics and answer vital questions about the diversity of ways PTMs alter RBP function.

As mass spectroscopy methodologies improve and become more accessible, protein modifications are mapped with high accuracy and with little cost [8]. Cataloguing modifications is the first step in order to understand how they participate in RBP function, and how RBP function is regulated by a certain pathway. Therefore, our intention in this chapter is far from reporting all the identified PTMs in translation control of RBPs. Instead, we will focus on some of the most studied cases hoping that they will serve as examples and predictions of what we may expected from the other members of this protein family.

There are several proteins or families of proteins discussed below that may have diverse roles that are not fully explored. As these are involved in several stages of RNA maturation, it is appropriate to first introduce them (Fig. 12.2):

HNRNPs are a diverse family of RNA-binding proteins for which post-translational modifications were discovered in the first descriptions of the proteins [9]. PTMs regulate the ability for HNRNP members to effect splicing changes and control the localization of RNA, as well as regulate translation, which will be discussed later in this chapter.

SR proteins were originally discovered for their role as splicing activators, but this view has been complicated by the nuances of effects due to PTMs [10, 11]. An abundance of literature points to the importance of PTMs in the activity of this family of proteins. Classically, SR proteins are modified by SR-protein kinases (SRPK1 and SRPK2) [12] and CDC2-like kinases (CLK1-4) [13–15]. The activity of these kinases is regulated through cell cycle, through development, and in response to cellular stresses like heat shock [10, 15, 16]. For several of the "classical" SR-proteins, including SRSF1 (aka Asf/Sf2) and SRSF2 (aka Sc35), phosphorylation induces changes in the intranuclear distribution of these phosphoproteins, causing release from nuclear speckles [17].


Protein domains
RRM
KH 1
KH 3
KH 2
B30.2/SPRY
CCHC-type
RRM 4
SAP
RRM 1
RRM 2
RRM 3
Modification type
Phosphorylation
Acetylation
Methylation
Citrulline
SRSF12
SRSF11
SRSF3
SRSF2
HNRNPCL1
HNRNPC
SRSF10
SRSF7
HNRNPUL1
HNRNPUL2
HNRNPU
HNRNPK
HNRNPLL
ELAVL1
ELAVL3
ELAVL2
HNRNPH2
HNRNPR
HNRNPF
HNRNPM
HNRNPH1
ELAVL4
HNRNPL
HNRNPAB
HNRNPDL
HNRNPA1L2
HNRNPD
SRSF6
SRSF5
SRSF9
HNRNPA0
SRSF4
SRSF1
HNRNPA3
HNRNPH3
HNRNPA2B1
HNRNPA1

ELAV (aka Hu) proteins are exemplified by a well-studied member of the family, ELAVL1. ELAVL1 (aka HuR, human antigen R) has 2 N-terminal RRM domains followed by a hinge region that contains a nucleo/cytoplasmic shuttling domain (HNS), and another C-terminal RRM domain. It recognizes and binds to Adenylate-Urydinilate rich elements (AREs) present in 3′UTR and/or 5′-UTR of transcripts in the nucleus and regulates their splicing, processing, nuclear export, localization, half-life and translation [18, 19]. Several features of its behavior are partially explained by PTMs. ELAVL1 has a dual effect on translation, being capable of activating translation of certain mRNAs (hypoxia-inducible factor (HIF)-1α, p53, prothymosin-α, MKP-1, cytochrome c, heme oxygenase-1, and cationic amino acid transporter 1 [20–25]) and inhibiting others (IGFR, p27, Wnt5a e c-Myc, [25–30] for instance) with altered RNA affinity dependent upon phosphorylation. Phosphorylation on the hinge domain affects ELAVL1 localization [31], as do several phosphorylation sites on the RRM domains. There is evidence that RRM3 domain phosphorylation modulates dimerization inducing higher substrate affinity and altered protein localization.

2 PTM-Mediated Regulation of Pre-mRNA Processing

2.1 Transcription

The integration of signaling cascades in the control of RNA begins with RNA polymerase. Synthesis, 5′ capping and 3′ polyadenylation of almost all protein-coding transcripts is orchestrated by phosphorylation of the C-terminal domain (CTD) of DNA-directed RNA-polymerase 2 (POLR2A and homologs B-M) [32]. CTD hyperphosphorylation by carboxy-terminal kinases induces the recruitment of capping enzymes [33, 34] and 3′ end cleavage and polyadenylation factors [35]. Phosphorylation of the CTD of POLR1* also regulates transcriptional activity of ribosomal RNAs [36]. In fact, there is an extensive body of literature devoted to the consequence of RNA-polymerase CTD phosphorylation and interested readers should read [32, 37] for a more comprehensive review of the effectors and effects of this specific PTM target. Further, acetylation by p300, indicated by a p300 dose-dependent shift in POLR2 CTD molecular weight, seems to be involved in transcription initiation or early transcription elongation of growth-factor induced genes [38]. The production of specific species of RNA that bind FUS nucleates the formation of nuclear FUS aggregates [39]. POLR2 CTD phosphorylation is reinforced by

Fig. 12.2 Large families of RBPs and their modifications. Proteins are shown as rectangles along rows. Functional domains of the protein are labelled and the proteins are sorted to group proteins with similar domains. PTMs listed in the Uniprot database are depicted as square shapes on each gene. Each vertical line is 50 amino acid residues. (Note: The QR code in the bottom-right links to an interactive version of this figure where references can also be visualized. Web link: here: https://rawgit.com/mlovci/12365bcafbef4a32d35a/raw/f781d50b3fd96ef83d019baf9e7984374420fdc6/Figure%25202.html)

these FUS aggregates [40]. Finally, FUS N-terminal phosphorylation by DNA protein-kinases (DNA-PK) removes FUS granules from the nucleus causing them to lose the potential to directly regulate POLR2 [41]. Thus, complex signaling and feedback control the activity of POLR2 through regulation of RBP PTM state.

Some HNRNPK PTMs serve to gate cell division checkpoints in response to DNA damage sensing. The activity of p53 tumor suppressor targets is tied to DNA damage-induced PTMs on the RNA-binding protein HNRNPK, which alter p53-HNRNPK protein-protein affinities. Methylation, phosphorylation and sumoylation of HNRNPK all regulate the p53-dependent cell cycle checkpoint [42–44]. Several signalling pathways converge to alter the function and stability of HNRNPK in response to DNA damage, including reduced expression of the E3 ligase MDM2 that targets HNRNPK to proteasomal destruction [45].

2.2 Splicing

Splicing is the RNA-catalyzed concatenation of exons that requires several protein scaffolds for which PTM state can control outcomes. Splicing in the nucleus is controlled by upstream signalling for DNA damage and cell cycle [46]. Indeed, it is closely tied to transcription and PTM state of histone proteins. For splicing components to mature, SMN complexes interact with U snRNAs and sm proteins to form snRNPs. This spliceosome formation occurs at Cajal bodies and requires the SMN complex. SMN components are localized, in part, by phosphorylation of the GEMIN proteins and deficiencies in this are linked to serious defects in intron recognition [47–49]. During spliceosome assembly, the targeted PTM of specific residues of snRNP must be required as both kinases and phosphatases are required for spliceosome assembly [50, 51].

S20 phosphorylation of SRSF1 initiates spliceosome assembly at intronic splice sites and is required for pre-mRNA processing fidelity [52]. This was reported to be regulated by the KIS kinase and important for bridging SRSF1 and U2AF2 in ternary SRSF1/U2AF2/RNA complexes [53]. Recent work shows with X-ray crystallography exactly the conformational shifts involved with phosphorylation of SRSF1 and reveals that only the phosphorylated version of SRSF1 can interact with U2AF65 [54]. SRSF10 (aka SRp38), is normally an unphosphorylated splicing repressor, but switches to a sequence-specific splicing activator when it is phosphorylated [17, 55, 56], presumably by inducing formation of spliceosomal complex A along with S100 [57].

2.3 Alternative Splicing

Alternative splicing (AS) is the regulated process of selective inclusion of specific exons into processed mRNA transcripts at specific stages of development or in response to external stimulation. Alternative splicing results from inefficient

recognition by the spliceosome or competition among 3′ splice-sites for ligation to 5′ splice-sites. Splicing factors are regulated through signalling cascades to either activate or repress splicing in certain environmental conditions. These include pathways that recognize extracellular signals like EGF, Wnt, insulin, cytokines and heat stress [16, 55, 58, 59].

Beside sub-nuclear localization and interactions with snRNPs, phosphorylation of SR-proteins has been shown to cause shuttling between the nucleus to the cytoplasm, usually resulting in the loss of inclusion of their splicing targets [60–62]. Proline-directed SRSF1 phosphorylation causes conformational shifts that affect enzymatic activity of the protein [63]. Lines of evidence implicating non-phosphorylation PTMs like ubiquitination and acetylation are less common but do exist, and are associated with regulation of protein turnover [64, 65]. In the case of acetylation, lysine-acetylated SRSF2 proteins by KAT5 (aka Tip60) are more likely to be subject to degradation and HDAC6-mediated deacetylation causes SRSF2 accumulation. However, KAT5-mediated PTM is accompanied by concomitant acetylation of SRPK1 and SRPK2 that causes these kinases to be excluded from the nucleus, thus the accumulated SR proteins are not actively regulating splicing in these cells [64]. Ubiquitination of SRSF1 was shown to be increased in activated T-cells, causing proteasomal destruction of the protein, but the E3 ligase that mediates this PTM is not yet known [65]. Development of small-molecule inhibitors of these SR-related PTMs has been the focus of recent research with potential applications in treatment of cancers like metastatic melanoma [66, 67].

HNRNPL S52 phosphorylation mediates signal integration via the PI3K/AKT pathway. Phosphorylated HNRNPL, but not non-phosphorylated HNRNPL out-competes HNRNPU for binding at a *cis*-element. RNA-binding assays with an antibody specific for S52-phosphorylated HNRNPL shows that when HNRNPL is phosphorylated, it is associates with RNA while HNRNPU binding is diminished leading to exclusion of a pro-apoptotic caspase-9 exon; HNRNPU phosphorylation alone did not account for this change [68]. Similarly, upon neuron depolarization CAMK4 kinase activation causes phosphorylation of HNRNPL. This S513 phosphorylation increases HNRNPL affinity for CaM-kinase responsive RNA elements, out-competing assembling spliceosomal components and inhibiting exon inclusion [69]. Data obtained with a methylation-sensitive antibody indicates that PRMT1 causes constitutive methylation on HNRNPU, but the authors did not observe methylation-dependent localization shifts and could not discern a regulated function for the methylation of this protein [70].

Splicing of the stress-induced isoform of the TRA2B transcript by ELAVL1 is accomplished only when nuclear-localized ELAVL1 is phosphorylated downstream of Chk2 and p38-MAPK at positions S88 and T118 [71]. These phosphorylated residues increase ELAVL1's affinity for an intronic binding site near an exon that causes an in-frame stop-codon, in turn causing higher levels of exon inclusion and subsequent nonsense-mediated decay of the TRA2B transcript.

KHDRBS1 (aka Sam68), a member of the STAR family of RNA-binding proteins, stands out in that it has reports of multiple classes of PTMs modify its RNA regulatory activity. Phosphorylation or acetylation increases its affinity for

RNA and splicing regulatory activity as shown with point-mutants and chemical small-molecule inhibitors of phosphorylation [59, 72–75] while methylation decreases affinity for poly-(U) targets [76]. Mutational studies showed that S58, T71 and T84 phosphorylation were required for splicing activation and authors note ATP-gS, a phosphotase-resistant ATP analog, was necessary to observe this effect in *in vitro* splicing assays [59]. In addition, tyrosine phosphorylation may influence the ability of KHDRBS1 to effectively form dimers, which are required for splicing-regulatory activity [77]. SUMOylation has also been reported on KHDRBS1 but not with an RNA-regulatory effect [78].

These are just a few of the hundreds of RNA-binding protein PTMs have clear roles in regulating downstream splicing. For example, RBFOX1 (by WNK3) and RBFOX2 (by PRKCA/B) are shown to be shuttled out of the nucleus and degraded, respectively, by phosphorylation; thus, excluding these proteins from regulating their target exons [79, 80]. TRA2B has a reduced affinity for the mRNA that encodes the TRA2B protein when it is phosphorylated by CLK2 [81]. CELF1 phosphorylation downstream of Akt signaling causes changes in subcellular CELF1 distributions, affecting splicing and translational control (reviewed in detail in [82]). Hyperphosphorylation of CELF1 by PKCA/B/C was shown to be downstream of accumulation of toxic DM1 repetitive RNA in myotonic dystrophy and important for proper splicing regulation [83].

2.4 mRNA 5′ G-Capping and Decapping

RNA 5′ 7-methyl guanosine capping by RNA guanlyltransferase, which protects RNAs from 5′ exonucleases, promotes translation and nuclear export, is tied to the phosphorylation state of RNA polymerase II CTD and this function is evolutionarily conserved to yeast [33, 35]. Decapping conversely is the first step of RNA decay and inhibits translation initiation. Decapping enzymes 1 and 2 in mammals are subject to rapid decay by ubiquitination and subsequent proteasomal degradation; thus, leading to longer RNA half-lives in general, in this case shown for a selection of targets that are subject to AU-rich element-mediated decay [84].

2.5 RNA Editing

ADAR protein levels, and consequently the extent of adenosine-to-inosine editing, are linked to the PTM state of these proteins. SUMO modification of ADAR1 at a lysine residue causes reduced editing efficacy *in vitro* and *in vivo* [85]. ADAR2 levels are decreased when phosphorylated by c-Jun kinase, resulting in reduced ADAR2-mediated A-to-I editing in pancreas [86].

3 PTM Regulation of Subcellular Localization

RBP PTMs commonly affect the ability for RBPs to move among cellular compartments. Thus, by virtue of their binding to RNA, RBPs regulate RNA localization based on their PTM state. In general, phosphorylation that affects RBP location also affects the set of bound RNAs, as may be expected since the availability of particular RNA species is not uniform across cells.

3.1 Nuclear/Cytoplasmic Shuttling

A few HNRNP proteins are sorted into the nucleus based on their PTM state. Nichols and colleagues showed with tritiated S-Adenosyl methionine then immunostaining after PRMT1 knockdown and GST-PRMT1 pull-down that HNRNPA2 arginine methylation in the RGG domain by PRMT1 is responsible for nuclear localization which is required for its regulation of alternative splicing [87]. HNRNPA1 localization is also regulated by PTM, with phosphorylation causing nuclear exclusion in a process that is activated by cellular stressors [88]. HNRNPQ has roles in splicing and mRNA stability and its localization is controlled by PRMT1-mediated methylation [89]; this may be important for controlling stability of RNA targets.

ELAVL1 is modified by an ubiquitin-like protein called NEDD8 on K313, K326 by MDM2 [90]. NEDD8 has 60 % homology with ubiquitin and its classical substrates are the cullin subunits of SCF ubiquitin E3 ligases [91]. Recently, it has been shown that MDM2 can associate with Ubc12 (the NEDD8 conjugating enzyme) and act as a NEDD8 ligase for p53 [92]. In the case of ELAVL1, neddylation promotes nuclear localization and inhibition of degradation [90].

Several kinases have been shown to be able to phosphorylate ELAVL1 and modulate its subcellular localization. For instance in the RRM domains: T118 phosphorylation by Chk2 or p38-MAPK [93, 94]; S158 phosphorylation by PRKCA and S318 phosphorylation by PRKCD [95, 96]. The hinge region (residues 186–244) is a hotspot for phosphorylation. Modifications on the hinge region affect ELAVL1 nucleo/cytoplasmic localization. Phosphorylation at S202 by CDK1 or CDK5, phosphorylation at S221 by PKC family members (PRKCA, PRKCD) and S242 phosphorylation by an unknown kinase all promote nuclear retention of the protein [31].

3.2 RNA Granules, P-Bodies and Nuclear Speckles

RNA-granules, processing bodies and nuclear speckles are functionally different aggregations of proteins and RNA that have modified activity and membership due to regulated changes in PTM state.

ELAVL1 phosphorylation on the hinge region outside of the HNS on Y200 by JAK3 inhibits ELAV1's localization to stress granules upon arsenite stress, leading to accelerated degradation of some of its mRNA targets (e.g. SIRT1 and VHL), but it is unclear whether mRNAs are bound to ELAVL1 during the transition to stress granules [97].

TARDBP (aka TDP-43) is acetylated at K145, K192 by CREBBP (Creb-binding protein). Based on crystal structure mapping of acetylated side-chains, the conformation of TARDBP RRM may shift and alter its ability to bind to RNA. Using glutamine to mimic acetylated lysine and forced acetylation by CREBBP, Cohen and colleagues show that acetylated lysine on TARDBP reduces RNA-binding and results in aggregation of TARDBP into cytoplasmic inclusions. When not bound to RNAs, TARDBP exits the nucleus, joins RNA granules and is phosphorylated at S410 [98], perhaps by CSK1 (casein kinase 1) [99] or TTBK1/2 (Tau tubulin kinases 1 & 2) [100]. This may represent coordinated handoff between post-translational modifiers to place TARDBP in granules. While it is possible to prevent neurodegeneration by blocking TARDBP phosphorylation at S409/S410 [101] or by preventing acetylation, it is not clear how these mechanisms interact to cause RNA granules. Further, there are other modifications that will certainly need to be considered, including ubiquitination, which likely follows aggregation and precedes proteassomal degradation [102]. These processes are of biomedical importance because phosphorylated TARDBP is a hallmark of cytoplasmic inclusions in Amyotrophic Lateral Sclerosis (ALS) [103] and mutations that are predicted to increase TARDBP phosphorylation are linked to ALS [104].

3.3 Exosome

Exosome loading of RNAs including microRNAs (miRNAs) and longer classes of noncoding RNA are controlled with PTMs of a few proteins. RBM7 phosphorylation downstream of the MAPK-mediated stress response sorts ncRNA in the nucleus to exosomes [105]. Exosome targeting for certain unprocessed miRNAs is similarly HNRNPA2B1 sumoylation-dependent [106]. KHSRP is an RBP phosphorylated through ATM and PI3K/ATM kinases in response to DNA damage that guides RNAs to the exosome, where they are targeted for destruction [107, 108]. Indeed exosome destruction of several RNAs is coordinated through signal integration/kinase activation of several proteins including RBM7, KHSRP, TTP and others [109].

4 PTM Regulation of Translation

Some cytoplasmic RBPs modulate protein output by either contributing to initiation, elongation or termination of translation of their mRNA substrates, thus having a large effect via downstream cellular processes coded by targeted mRNAs. Below,

we describe a few translation-regulatory RBPs, which are modified by phosphorylation, methylation, ubiquitination and oxidation.

Besides modulating subcellular localization and splicing, phosphorylation of the ELAVL1, as discussed above, PTM in ELAVL1 RRM domains can also modulate substrate affinity. For instance S38 and S100 phosphorylation by Chk2 can modulate mRNA substrate recognition. Interestingly, S100 phosphorylation seems to define the selectivity in ELAVL1 targeting. Thus, while phosphorylation induces release of Sirt1 mRNA and subsequent destabilization of Sirt1 mRNA, [93], the opposite is observed where S100 phosphorylation has been reported to increase affinity of ELAVL1 for Occludin mRNA (increasing its translation efficiency, [110]). This dual effect suggests the interesting possibility that phosphorylation may be used as a way for ELAVL1 to discriminate its different substrates and integrate dynamic signaling cues to translation decisions. It is unclear to what extent ELAVL1's promiscuity in target selection is subject to this S100 phosphorylation event or other PTMs.

At least methylation and phosphorylation modulate FMR1 protein (aka FMRP) function. Absence of FMR1 expression in neurons leads to developmental abnormalities, such as immature, thin and highly branched dendritic spines. FMR1 has two Agenet domains followed by a NLS (nuclear localization signal) sequence close to its N-terminal, two KH domains (HNRNPK-homology domain) followed by a NES (nuclear export signal) signal in its middle and a RGG domain (arginine-glycine-glycine domain). The KH domain and RGG domains bind RNA. Phosphorylated FMR1 is associated with stalled polyribosome complexes. FMR1 forms a translation-inhibitory complex with the target mRNA and Cytoplasmic FMR1 Interacting Protein (CYFIP1). This complex binds translation protein EIF4E, thereby inhibiting its interaction with EIF4G. Phosphorylation is in fact necessary for FMR1 to carry out its roles in developmental timing [111]. FMR1 may also bind directly to the ribosome in polysomes to inhibit translation elongation [112]. Indeed phosphorylation status of FMR1 has also been proposed to regulate its association with translating ribosomes and stalled ribosomes. Arginine methylation of the FMR1 RGG domain by PRMT1 has being proposed to inhibit its ability to recognize target mRNAs and its assembly in translation initiation inhibitory complexes [113–115].

CPEB1 contains a PEST sequence, two conserved RRM domains and a c-terminal ZNF-domain [116]. This RBP modulates translation of target mRNAs which contain a Cytoplasmic polyadenylation element (CPE) in their 3′UTR [117]. CEPB1 binds to CPE mRNA substrates, keeping them in a translation inhibited state. When CEBP1 is phosphorylated on S174 (outside of the RRM domains) by Aurora A kinase [118–120], it recruits CPSF [119, 121] and induces polyadenylation of the mRNA, greatly inducing their expression. CEBP1 can also be phosphorylated sequentially by Cdc2 on T125 and multiple Serine residues, which recruits Plx1 that phosphorylate S191 on the PEST sequence. Once PEST is phosphorylated, the hyperphosphorylated CPEB1 is recognized by the SCFb-TrCP E3 ubiquitin ligase complex, and polyubiquitinated leading to its proteassomal degradation [122]. CPEB1 plays important roles in the development of Xenopus oocyte and synapse formation/long-term potentiation [123, 124].

5 PTM Regulation of RNA Stability and Destruction

5.1 miRNA Related Repression

miRNAs are produced through a two-stage double-stranded RNA cleavage and processing. The miRNA pathway can either affect mRNA stability or translational output through miRNA:mRNA base-pairing mediated in a RNA-protein complex called RNA-induced silencing complex (RISC). Canonical miRNAs originate from long primary mRNA transcripts, which are initially processed to miRNA precursors by the nuclear microprocessor complex in animals. The ribonuclease DROSHA and its RBP partner DGCR8 (DiGeorge syndrome critical region gene 8) are two key components of this complex (reviewed in [125]). PTM regulates this step in miRNA maturation, evidenced by co-immunoprecipitation of DROSHA and DGCR8 with components of the class I of Histone deacetylases (HDAC) [126]. In addition, the overexpression of HDAC1 in HEK293 cells results in increased affinity for primary miRNAs and higher mature miRNA availability without increasing the in corresponding primary miRNA's expression levels [126]. This was attributed to deacetylation of the DGCR8 RNA-binding domain and increased affinity to primary miRNAs [126]. Other studies have also proposed that Drosha and DGCR8 can be stabilized by phosphorylation, and it was shown that anti-MAPK/CDK substrate antibodies recognized immunopurified DGCR8 [127].

Microprocessor components are not the only proteins targeted by PTM for regulation of nuclear miRNA processing. Certain factors that regulate miRNA biogenesis are altered by PTMs. In early differentiation, LIN28 is expressed and binds to the let-7 primary transcript, but acetylation by PCAF and ubiquitination by TRIM71 causes destruction of LIN28 leading to de-repressed let-7 processing and allowing cells to progress through differentiation [128, 129]. The E3 ligase TRIM65 represses RISC assembly by targeting TNRC6 (aka GW182) for destruction [130].

After miRNA precursors are loaded into the RISC and exported from the nucleus, DICER1, an RNAseIII, is responsible for recognizing the hairpin precursor sequences and processes them to mature miRNAs. FMR1 has been shown to interact with DICER1, argonaute 2 (AGO2) and specific miRNAs. Phosphorylation holds FMR1 in association with its targets and prevents translation, perhaps by AGO2 interactions with targets. While at the same time, phosphorylation of FMRP inhibits association with DICER and reduces DICER activity [131, 132].

Argonaute proteins facilitate the interactions between the 22nt long microRNAs and target mRNAs. Several signaling pathways converge on these AGO proteins to control their activity in various cellular contexts [133, 134]. AGO protein is phosphorylated by p38-MAPK under cellular stress treatments like sodium arsenite, causing it to localize to processing bodies [135]. Certain AGO family members are preferentially subject to hydroxylation, stabilizing these proteins and potentiating the effect of miRNAs in hypoxia [134, 136].

One study suggests that phosphorylation at S499 (in the RGG domain) of FMR1 modulates translation of its target mRNA and via AGO2 [132]. Activation of mGluR

pathway leads to dephosphorylation of FMR1, followed by disassembly of its associated translation inhibitory complex and induction of PSD-95 translation/protein expression. It has being proposed that RPS6KB1 (aka S6K1) is the kinase that phosphorylates FMR1 and PP2A is the phosphatase that dephosphorylates it in the mGluR pathway [137, 138]. A separate study showed that HOXB8 mRNA is subject to the same phospho-FMR1/AGO2/miR-196a inhibitory complex [139]. This mechanism may be a way to induce gene expression of several FMR1 targets during long-term synaptic depression as depolarization leads to PP2A activation [138].

5.2 RNA Decay

QKI, a STAR-family RBP, is an essential regulator of myelination in oligodendrocytes. At early stages of development, QKI binds and stabilizes MBP RNA. C-terminal phosphorylation by Src-Protein Tyrosine Kinases (PTK) at Y285, Y288, Y290, Y292 and Y303 decreases QKI's affinity for MBP (myelin basic protein) mRNA [140]. As src-PTK activity is reduced in early myelin development, mRNA can associate with QKI and accumulate. Indeed QKI sits downstream of several developmentally and disease-relevant pathways and understanding how PTMs affect its function will be an important goal for future studies ([140], reviewed in [141]).

ELAVL1 ubiquitination is related to the stability of its targets. ELAVL1 K48-linked ubiquitination on K182 by an unknown ubiquitin ligase promotes its proteasomal degradation [142]. However, K29-linked ubiquitination on ELAVL1 K313/K326 is reported to be a signal for protein-RNA complex disassembly. These modifications induce release of some ELAVL1 substrates (p21, MKP-1, and SIRT1 mRNAs) from ELAVL1, through recruitment of the p97–UBXD8 complex, leading to their destabilization [143]. Localization is also changed upon methylation of the hinge region by CARM1 (co-activator-associated arginine methyltransferase 1) at R217 [144]. Although the functional consequence of this modification is not completely understood, it has been shown to enhance ELAVL1's ability to regulate turnover of some of its substrate mRNAs (TNF-alpha, cyclin A, cyclin B1, c-fos, SIRT1, and p16) [145].

6 Conclusion

As the effects of PTMs are broad and unpredictable, careful follow-up on the dynamic changes in protein function are necessary. With regard to RBP function, the essential question is whether a particular PTM will affect many of the RBPs in ways we have listed above, including:

1. RNA-binding ability (i.e. QKI)
2. Protein complex formation (i.e. SRSF1/U2AF65)
3. Subcellular localization (i.e. CELF1, KDHRBS1)

(a) Are RNAs bound during the transit?
(b) What mechanisms drive RBP motility?

4. Enzymatic activity of the RBP (i.e. ADAR)
5. Initiation of RBP for destruction (i.e. LIN28)

Although the downstream consequences of these processes are varied and will be intensely studied, these are the basic features of RBP functions affected by PTMs. Answers to these simple questions for the library of RBP PTMs will be critical for accurate modeling of the effect of context and signal integration into the mRNP code. No doubt, advances in methods for probing protein structure will glean insight into the potential roles for PTMs on RBP function and provide a guide to prioritize the search for PTMs that have an impact on RNA maturation.

It should be noted again that this brief chapter is in no way a complete summary of the catalog of PTMs on RBPs. In the interest of space we had to restrict our discussions to the most well characterized examples and have left out many examples that may be relevant for basic biology or disease. These include rare post-translational modifications like nitration, which has evidence for affecting HNRNPA2B1 proteins [146], prolyl isomerization of POL2R [147], myristoylation which affects the axonal distribution of FXR2 [148] and PARylation which can globally repress the miRNA pathway in stress [149]. Several reviews cited herein have approached the relevance of PTMs in a particular pathway, family of genes, biological process or disease. Interested readers should follow this text with a thorough examination of these and the associated primary literature. There is a great deal still unknown about the cumulative and cross-regulatory effects that each PTM on each RBP holds. If the history of DNA-binding proteins and histone modifications is any indicator, this will be an area ripe for discovery and will advance basic biology and drug development.

Acknowledgments We would like to thank members of the Massirer and Bengston labs for their critical reading of this manuscript. We also would like to acknowledge the many works not included herein that have contributed to the understanding of the role of RNA-binding protein post-translational modifications.

References

1. Castello A, Fischer B, Hentze MW, Preiss T (2013) RNA-binding proteins in Mendelian disease. Trends Genet 29:318–327
2. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
3. Kapeli K, Yeo GW (2012) Genome-wide approaches to dissect the roles of RNA-binding proteins in translational control: implications for neurological diseases. Front Neurosci 6:144
4. Lukong KE, Chang KW, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends Genet 24:416–425
5. Nussbacher JK, Batra R, Lagier-Tourenne C, Yeo GW (2015) RNA-binding proteins in neurodegeneration: Seq and you shall receive. Trends Neurosci 38:226–236

6. Konig H, Ponta H, Herrlich P (1998) Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. EMBO J 17:2904–2913
7. Michel MA, Elliott PR, Swatek KN, Simicek M, Pruneda JN, Wagstaff JL, Freund SM, Komander D (2015) Assembly and specific recognition of k29- and k33-linked polyubiquitin. Mol Cell 58:95–109
8. Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, Gygi SP (2008) A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A 105:10762–10767
9. Soulard M, Della Valle V, Siomi MC, Pinol-Roma S, Codogno P, Bauvy C, Bellini M, Lacroix JC, Monod G, Dreyfuss G et al (1993) hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. Nucleic Acids Res 21:4210–4217
10. Sanford JR, Bruzik JP (1999) Developmental regulation of SR protein phosphorylation and activity. Genes Dev 13:1513–1518
11. Xiang-Yang Z, Pingping W, Joonhee H, Michael GR, Xiang-Dong F (2009) SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Mol Cell 35(1):1–10
12. Wang HY, Lin W, Dyck JA, Yeakley JM, Songyang Z, Cantley LC, Fu XD (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J Cell Biol 140:737–750
13. Aubol B, Plocinik R, Hagopian J, Ma C, McGlone M, Bandyopadhyay R, Fu X, Adams J (2013) Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. J Mol Biol 425:2894–2909
14. Aubol B, Plocinik R, Keshwani M, McGlone M, Hagopian J, Ghosh G, Fu X, Adams J (2014) N-terminus of the protein kinase CLK1 induces SR protein hyperphosphorylation. Biochem J 462:143–152
15. Ninomiya K, Kataoka N, Hagiwara M (2011) Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. J Cell Biol 195:27–40
16. Shin C, Manley JL (2002) The SR protein SRp38 represses splicing in M phase cells. Cell 111:407–417
17. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J 15:265–275
18. Hinman MN, Lou H (2008) Diverse molecular functions of Hu proteins. Cell Mol Life Sci 65:3168–3181
19. López de Silanes I, Zhan M, Lal A, Yang X, Gorospe M (2004) Identification of a target RNA motif for RNA-binding protein HuR. Proc Natl Acad Sci U S A 101:2987–2992
20. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell 125:1111–1124
21. Kawai T, Lal A, Yang X, Galban S, Mazan-Mamczarz K, Gorospe M (2006) Translational control of cytochrome c by RNA-binding proteins TIA-1 and HuR. Mol Cell Biol 26:3295–3307
22. Kuwano Y, Kim HH, Abdelmohsen K, Pullmann R, Martindale JL, Yang X, Gorospe M (2008) MKP-1 mRNA stabilization and translational control by RNA-binding proteins HuR and NF90. Mol Cell Biol 28:4562–4575
23. Kuwano Y, Rabinovic A, Srikantan S, Gorospe M, Demple B (2009) Analysis of nitric oxide-stabilized mRNAs in human fibroblasts reveals HuR-dependent heme oxygenase 1 upregulation. Mol Cell Biol 29:2622–2635
24. Lal A, Kawai T, Yang X, Mazan-Mamczarz K, Gorospe M (2005) Antiapoptotic function of RNA-binding protein HuR effected through prothymosin alpha. EMBO J 24:1852–1862
25. Mazan-Mamczarz K, Galbán S, López de Silanes I, Martindale JL, Atasoy U, Keene JD, Gorospe M (2003) RNA-binding protein HuR enhances p53 translation in response to ultraviolet light irradiation. Proc Natl Acad Sci U S A 100:8354–8359
26. Abdelmohsen K, Kuwano Y, Kim HH, Gorospe M (2008) Posttranscriptional gene regulation by RNA-binding proteins during oxidative stress: implications for cellular senescence. Biol Chem 389:243–255

27. Kim HH, Kuwano Y, Srikantan S, Lee EK, Martindale JL, Gorospe M (2009) HuR recruits let-7/RISC to repress c-Myc expression. Genes Dev 23:1743–1748
28. Kullmann M, Göpfert U, Siewe B, Hengst L (2002) ELAV/Hu proteins inhibit p27 translation via an IRES element in the p27 5′UTR. Genes Dev 16:3087–3099
29. Leandersson K, Riesbeck K, Andersson T (2006) Wnt-5a mRNA translation is suppressed by the Elav-like protein HuR in human breast epithelial cells. Nucleic Acids Res 34:3988–3999
30. Meng Z, King PH, Nabors LB, Jackson NL, Chen C-Y, Emanuel PD, Blume SW (2005) The ELAV RNA-stability factor HuR binds the 5′-untranslated region of the human IGF-IR transcript and differentially represses cap-dependent and IRES-mediated translation. Nucleic Acids Res 33:2962–2979
31. Kim HH, Yang X, Kuwano Y, Gorospe M (2008) Modification at HuR(S242) alters HuR localization and proliferative influence. Cell Cycle 7:3371–3377
32. Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev 20:2922–2936
33. Cho EJ, Takagi T, Moore CR, Buratowski S (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev 11:3319–3326
34. McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL (1997) 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev 11:3306–3318
35. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385:357–361
36. Bouchoux C, Hautbergue G, Grenetier S, Carles C, Riva M, Goguel V (2003) CTD kinase I is involved in RNA polymerase I transcription. Nucleic Acids Res 32:5851–5860
37. Hsin J-PP, Xiang K, Manley JL (2014) Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Mol Cell Biol 34:2488–2498
38. Schröder S, Herker E, Itzen F, He D, Thomas S, Gilchrist DA, Kaehlcke K, Cho S, Pollard KS, Capra JA et al (2013) Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Mol Cell 52:314–324
39. Schwartz JC, Wang X, Podell ER, Cech TR (2013) RNA seeds higher-order assembly of FUS protein. Cell Rep 5:918–925
40. Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell 155:1049–1060
41. Deng Q, Holler CJ, Taylor G, Hudson KF, Watkins W, Gearing M, Ito D, Murray ME, Dickson DW, Seyfried NT et al (2014) FUS is phosphorylated by DNA-PK and accumulates in the cytoplasm after DNA damage. J Neurosci 34:7802–7813
42. Lee SW, Lee MH, Park JH, Kang SH, Yoo HM, Ka SH, Oh YM, Jeon YJ, Chung CH (2012) SUMOylation of hnRNP-K is required for p53-mediated cell-cycle arrest in response to DNA damage. EMBO J 31:4441–4452
43. Pelisch F, Pozzi B, Risso G, Muñoz MJ, Srebrow A (2012) DNA damage-induced heterogeneous nuclear ribonucleoprotein K sumoylation regulates p53 transcriptional activation. J Biol Chem 287:30789–30799
44. Yang JH, Chiou YY, Fu SL, Shih IY, Weng TH, Lin WJ, Lin CH (2014) Arginine methylation of hnRNPK negatively modulates apoptosis upon DNA damage through local regulation of phosphorylation. Nucleic Acids Res 42:9908–9924
45. Moumen A, Masterson P, O'Connor MJ, Jackson SP (2005) hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. Cell 123:1065–1078
46. Moore MJ, Wang Q, Kennedy CJ, Silver PA (2010) An alternative splicing network links cell-cycle control to apoptosis. Cell 142:625–636
47. Husedzinovic A, Oppermann F, Draeger-Meurer S, Chari A, Fischer U, Daub H, Gruss OJ (2014) Phosphoregulation of the human SMN complex. Eur J Cell Biol 93:106–117

48. Husedzinovic A, Neumann B, Reymann J, Draeger-Meurer S, Chari A, Erfle H, Fischer U, Gruss OJ (2015) The catalytically inactive tyrosine phosphatase HD-PTP/PTPN23 is a novel regulator of SMN complex localization. Mol Biol Cell 26:161–171
49. Renvoisé B, Quérol G, Verrier ER, Burlet P, Lefebvre S (2012) A role for protein phosphatase PP1γ in SMN complex formation and subnuclear localization to Cajal bodies. J Cell Sci 125:2862–2874
50. Kamoun M, Filali M, Murray MV, Awasthi S, Wadzinski BE (2013) Protein phosphatase 2A family members (PP2A and PP6) associate with U1 snRNP and the spliceosome during pre-mRNA splicing. Biochem Biophys Res Commun 440:306–311
51. Mathew R, Hartmuth K, Mohlmann S, Urlaub H, Ficner R, Luhrmann R (2008) Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat Struct Mol Biol 15:435–443
52. Wang X, Bruderer S, Rafi Z, Xue J, Milburn PJ, Krämer A, Robinson PJ (1999) Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. EMBO J 18:4549–4559
53. Manceau V, Swenson M, Le Caer J-PP, Sobel A, Kielkopf CL, Maucuer A (2006) Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. FEBS J 273:577–587
54. Wang W, Maucuer A, Gupta A, Manceau V, Thickman KR, Bauer WJ, Kennedy SD, Wedekind JE, Green MR, Kielkopf CL (2013) Structure of phosphorylated SF1 bound to U2AF(6)(5) in an essential splicing factor complex. Structure 21:197–208
55. Shin C, Feng Y, Manley JL (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature 427:553–558
56. Shin C, Kleiman FE, Manley JL (2005) Multiple properties of the splicing repressor SRp38 distinguish it from typical SR proteins. Mol Cell Biol 25:8334–8343
57. Feng Y, Chen M, Manley JL (2008) Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat Struct Mol Biol 15:1040–1048
58. Fu XD, Ares M Jr (2014) Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 15:689–701
59. Matter N, Herrlich P, Konig H (2002) Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature 420:691–695
60. Caceres J, Screaton G, Krainer A (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev 12:55–66
61. Lai MC, Tarn WY (2004) Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. J Biol Chem 279:31745–31749
62. Lai MC, Lin RI, Huang SY, Tsai CW, Tarn WY (2000) A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. J Biol Chem 275:7950–7957
63. Keshwani MM, Aubol BE, Fattet L, Ma CT, Qiu J, Jennings PA, Fu XD, Adams JA (2015) Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. Biochem J 466:311–322
64. Edmond V, Moysan E, Khochbin S, Matthias P, Brambilla C, Brambilla E, Gazzeri S, Eymin B (2011) Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. EMBO J 30:510–523
65. Moulton VR, Gillooly AR, Tsokos GC (2014) Ubiquitination regulates expression of the serine/arginine-rich splicing factor 1 (SRSF1) in normal and systemic lupus erythematosus (SLE) T cells. J Biol Chem 289:4126–4134
66. Gammons MV, Lucas R, Dean R, Coupland SE, Oltean S, Bates DO (2014) Targeting SRPK1 to control VEGF-mediated tumour angiogenesis in metastatic melanoma. Br J Cancer 111:477–485
67. Ohe K, Hagiwara M (2015) Modulation of alternative splicing with chemical compounds in new therapeutics for human diseases. ACS Chem Biol 10:914–924
68. Vu NT, Park MA, Shultz JC, Goehe RW, Hoeferlin LA, Shultz MD, Smith SA, Lynch KW, Chalfant CE (2013) hnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. J Biol Chem 288:8575–8584

69. Liu G, Razanau A, Hai Y, Yu J, Sohail M, Lobo VG, Chu J, Kung SK, Xie J (2012) A conserved serine of heterogeneous nuclear ribonucleoprotein L (hnRNP L) mediates depolarization-regulated alternative splicing of potassium channels. J Biol Chem 287:22709–22716
70. Herrmann F, Bossert M, Schwander A, Akgün E, Fackelmayer FO (2004) Arginine methylation of scaffold attachment factor A by heterogeneous nuclear ribonucleoprotein particle-associated PRMT1. J Biol Chem 279:48774–48779
71. Akaike Y, Masuda K, Kuwano Y, Nishida K, Kajita K, Kurokawa K, Satake Y, Shoda K, Imoto I, Rokutan K (2014) HuR regulates alternative splicing of the TRA2β gene in human colon cancer cells under oxidative stress. Mol Cell Biol 34:2857–2873
72. Babic I, Jakymiw A, Fujita D (2004) The RNA-binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA-binding activity. Oncogene 23:3781–3789
73. Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C (2007) The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J Cell Biol 176:929–939
74. Taylor SJ, Shalloway D (1994) An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. Nature 368:867–871
75. Wang L, Richard S, Shaw A (1995) P62 association with RNA is regulated by tyrosine phosphorylation. J Biol Chem 270(5):2010–2013
76. Rho J, Choi S, Jung C-RR, Im D-SS (2007) Arginine methylation of Sam68 and SLM proteins negatively regulates their poly(U) RNA-binding activity. Arch Biochem Biophys 466:49–57
77. Meyer NH, Tripsianes K, Vincendeau M, Madl T, Kateb F, Brack-Werner R, Sattler M (2010) Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. J Biol Chem 285:28893–28901
78. Babic I, Cherry E, Fujita DJ (2006) SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis. Oncogene 25:4955–4964
79. Lee AY, Chen W, Stippec S, Self J, Yang F, Ding X, Chen S, Juang YC, Cobb MH (2012) Protein kinase WNK3 regulates the neuronal splicing factor Fox-1. Proc Natl Acad Sci U S A 109:16841–16846
80. Verma SK, Deshmukh V, Liu P, Nutter CA, Espejo R, Hung M-LL, Wang G-SS, Yeo GW, Kuyumcu-Martinez MN (2013) Reactivation of fetal splicing programs in diabetic hearts is mediated by protein kinase C signaling. J Biol Chem 288:35372–35386
81. Stoilov P, Daoud R, Nayler O, Stamm S (2004) Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. Hum Mol Genet 13:509–524
82. Giudice J, Cooper T (2014) RNA-binding proteins in heart development. Adv Exp Med Biol 825:389–429
83. Kuyumcu-Martinez NM, Wang G-SS, Cooper TA (2007) Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Mol Cell 28:68–78
84. Erickson SL, Corpuz EO, Maloy JP, Fillman C, Webb K, Bennett EJ, Lykke-Andersen J (2015) Competition between decapping complex formation and ubiquitin-mediated proteasomal degradation controls human Dcp2 decapping activity. Mol Cell Biol 35:2144–2153
85. Desterro JM, Keegan LP, Jaffray E, Hay RT, O'Connell MA, Carmo-Fonseca M (2005) SUMO-1 modification alters ADAR1 editing activity. Mol Biol Cell 16:5115–5126
86. Yang L, Huang P, Li F, Zhao L, Zhang Y, Li S, Gan Z, Lin A, Li W, Liu Y (2012) c-Jun amino-terminal kinase-1 mediates glucose-responsive upregulation of the RNA editing enzyme ADAR2 in pancreatic beta-cells. PLoS One 7(11):e48611
87. Nichols RC, Wang XW, Tang J, Hamilton BJ, High FA, Herschman HR, Rigby WF (2000) The RGG domain in hnRNP A2 affects subcellular localization. Exp Cell Res 256:522–532
88. van der Houven van Oordt W, Diaz-Meco MT, Lozano J, Krainer AR, Moscat J, Cáceres JF (2000) The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. J Cell Biol 149:307–316
89. Passos DO, Quaresma AJ, Kobarg J (2006) The methylation of the C-terminal region of hnRNPQ (NSAP1) is important for its nuclear localization. Biochem Biophys Res Commun 346:517–525
90. Embade N, Fernández-Ramos D, Varela-Rey M, Beraza N, Sini M, Gutiérrez de Juan V, Woodhoo A, Martínez-López N, Rodríguez-Iruretagoyena B, Bustamante FJ et al (2012) Murine

double minute 2 regulates Hu antigen R stability in human liver and colon cancer through NEDDylation. Hepatology 55:1237–1248
91. Enchev RI, Schulman BA, Peter M (2015) Protein neddylation: beyond cullin-RING ligases. Nat Rev Mol Cell Biol 16:30–44
92. Batuello CN, Hauck PM, Gendron JM, Lehman JA, Mayo LD (2015) Src phosphorylation converts Mdm2 from a ubiquitinating to a neddylating E3 ligase. Proc Natl Acad Sci U S A 112:1749–1754
93. Abdelmohsen K, Pullmann R, Lal A, Kim HH, Galban S, Yang X, Blethrow JD, Walker M, Shubert J, Gillespie DA et al (2007) Phosphorylation of HuR by Chk2 regulates SIRT1 expression. Mol Cell 25:543–557
94. Lafarga V, Cuadrado A, Lopez de Silanes I, Bengoechea R, Fernandez-Capetillo O, Nebreda AR (2009) p38 Mitogen-activated protein kinase- and HuR-dependent stabilization of p21(Cip1) mRNA mediates the G(1)/S checkpoint. Mol Cell Biol 29:4341–4351
95. Doller A, Huwiler A, Müller R, Radeke HH, Pfeilschifter J, Eberhardt W (2007) Protein kinase C alpha-dependent phosphorylation of the mRNA-stabilizing factor HuR: implications for posttranscriptional regulation of cyclooxygenase-2. Mol Biol Cell 18:2137–2148
96. Doller A, Akool E-S, Huwiler A, Müller R, Radeke HH, Pfeilschifter J, Eberhardt W (2008) Posttranslational modification of the AU-rich element binding protein HuR by protein kinase Cdelta elicits angiotensin II-induced stabilization and nuclear export of cyclooxygenase 2 mRNA. Mol Cell Biol 28:2608–2625
97. Yoon J-H, Abdelmohsen K, Srikantan S, Guo R, Yang X, Martindale JL, Gorospe M (2014) Tyrosine phosphorylation of HuR by JAK3 triggers dissociation and degradation of HuR target mRNAs. Nucleic Acids Res 42:1196–1208
98. Cohen T, Hwang A, Restrepo C, Yuan C, Trojanowski J, Lee V (2015) An acetylation switch controls TDP-43 function and aggregation propensity. Nat Commun 6:5845
99. Kametani F, Nonaka T, Suzuki T, Arai T, Dohmae N, Akiyama H, Hasegawa M (2009) Identification of casein kinase-1 phosphorylation sites on TDP-43. Biochem Biophys Res Commun 382:405–409
100. Liachko NF, McMillan PJ, Strovas TJ, Loomis E, Greenup L, Murrell JR, Ghetti B, Raskind MA, Montine TJ, Bird TD et al (2014) The tau tubulin kinases TTBK1/2 promote accumulation of pathological TDP-43. PLoS Genet 10:e1004803
101. Liachko N, McMillan P, Guthrie C, Bird T, Leverenz J, Kraemer B (2013) CDC7 inhibition blocks pathological TDP-43 phosphorylation and neurodegeneration. Ann Neurol 74:39–52
102. Hans F, Fiesel FC, Strong JC, Jackel S, Rasse TM, Geisler S, Springer W, Schulz JB, Voigt A, Kahle PJ (2014) UBE2E ubiquitin-conjugating enzymes and ubiquitin isopeptidase Y regulate TDP-43 protein ubiquitination. J Biol Chem 289:19164–19179
103. Arai T, Hasegawa M, Nonoka T, Kametani F, Yamashita M, Hosokawa M, Niizato K, Tsuchiya K, Kobayashi Z, Ikeda K et al (2010) Phosphorylated and cleaved TDP-43 in ALS, FTLD and other neurodegenerative disorders and in cellular models of TDP-43 proteinopathy. Neuropathology 30:170–181
104. Kühnlein P, Sperfeld A-DD, Vanmassenhove B, Van Deerlin V, Lee VM, Trojanowski JQ, Kretzschmar HA, Ludolph AC, Neumann M (2008) Two German kindreds with familial amyotrophic lateral sclerosis due to TARDBP mutations. Arch Neurol 65:1185–1189
105. Tiedje C, Lubas M, Tehrani M, Menon MB, Ronkina N, Rousseau S, Cohen P, Kotlyarov A, Gaestel M (2015) p38MAPK/MK2-mediated phosphorylation of RBM7 regulates the human nuclear exosome targeting complex. RNA 21:262–278
106. Villarroya-Beltri C, Gutierrez-Vazquez C, Sanchez-Cabo F, Perez-Hernandez D, Vazquez J, Martin-Cofreces N, Martinez-Herrera DJ, Pascual-Montano A, Mittelbrunn M, Sanchez-Madrid F (2013) Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. Nat Commun 4:2980
107. Gherzi R, Trabucchi M, Ponassi M, Ruggiero T, Corte G, Moroni C, Chen CY, Khabar KS, Andersen JS, Briata P (2006) The RNA-binding protein KSRP promotes decay of beta-catenin mRNA and is inactivated by PI3K-AKT signaling. PLoS Biol 5:e5
108. Liu Y, Liu Q (2011) ATM signals miRNA biogenesis through KSRP. Mol Cell 41:367–368

109. Tiedje C, Holtmann H, Gaestel M (2014) The role of mammalian MAPK signaling in regulation of cytokine mRNA stability and translation. J Interferon Cytokine Res 34:220–232
110. Yu T-X, Wang P-Y, Rao JN, Zou T, Liu L, Xiao L, Gorospe M, Wang J-Y (2011) Chk2-dependent HuR phosphorylation regulates occludin mRNA translation and epithelial barrier function. Nucleic Acids Res 39:8472–8487
111. Say E, Tay H-GG, Zhao Z-S, Baskaran Y, Li R, Lim L, Manser E (2010) A functional requirement for PAK1 binding to the KH(2) domain of the fragile X protein-related FXR1. Mol Cell 38:236–249
112. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW et al (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 146:247–261
113. Dolzhanskaya N, Merz G, Aletta JM, Denman RB (2006) Methylation regulates the intracellular protein-protein and protein-RNA interactions of FMRP. J Cell Sci 119:1933–1946
114. Dolzhanskaya N, Merz G, Denman RB (2006) Alternative splicing modulates protein arginine methyltransferase-dependent methylation of fragile X syndrome mental retardation protein. Biochemistry 45:10385–10393
115. Stetler A, Winograd C, Sayegh J, Cheever A, Patton E, Zhang X, Clarke S, Ceman S (2006) Identification and characterization of the methyl arginines in the fragile X mental retardation protein Fmrp. Hum Mol Genet 15:87–96
116. Hake LE, Richter JD (1994) CPEB is a specificity factor that mediates cytoplasmic polyadenylation during Xenopus oocyte maturation. Cell 79:617–627
117. de Moor CH, Richter JD (1999) Cytoplasmic polyadenylation elements mediate masking and unmasking of cyclin B1 mRNA. EMBO J 18:2294–2303
118. Mendez R, Hake LE, Andresson T, Littlepage LE, Ruderman JV, Richter JD (2000) Phosphorylation of CPE binding factor by Eg2 regulates translation of c-mos mRNA. Nature 404:302–307
119. Mendez R, Murthy KG, Ryan K, Manley JL, Richter JD (2000) Phosphorylation of CPEB by Eg2 mediates the recruitment of CPSF into an active cytoplasmic polyadenylation complex. Mol Cell 6:1253–1259
120. Mendez R, Barnard D, Richter JD (2002) Differential mRNA translation and meiotic progression require Cdc2-mediated CPEB destruction. EMBO J 21:1833–1844
121. Dickson KS, Bilger A, Ballantyne S, Wickens MP (1999) The cleavage and polyadenylation specificity factor in Xenopus laevis oocytes is a cytoplasmic factor involved in regulated polyadenylation. Mol Cell Biol 19:5707–5717
122. Setoyama D, Yamashita M, Sagata N (2007) Mechanism of degradation of CPEB during Xenopus oocyte maturation. Proc Natl Acad Sci U S A 104:18001–18006
123. Di Nardo AA, Nedelec S, Trembleau A, Volovitch M, Prochiantz A, Montesinos ML (2007) Dendritic localization and activity-dependent translation of Engrailed1 transcription factor. Mol Cell Neurosci 35:230–236
124. Wu L, Wells D, Tay J, Mendis D, Abbott MA, Barnitt A, Quinlan E, Heynen A, Fallon JR, Richter JD (1998) CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. Neuron 21:1129–1139
125. Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15:509–524
126. Wada T, Kikuchi J, Furukawa Y (2012) Histone deacetylase 1 enhances microRNA processing via deacetylation of DGCR8. EMBO Rep 13:142–149
127. Herbert KM, Pimienta G, DeGregorio SJ, Alexandrov A, Steitz JA (2013) Phosphorylation of DGCR8 increases its intracellular stability and induces a progrowth miRNA profile. Cell Rep 5:1070–1081
128. Lee SH, Cho S, Kim MS, Choi K, Cho JY, Gwak HS, Kim YJ, Yoo H, Lee SH, Park JB et al (2014) The ubiquitin ligase human TRIM71 regulates let-7 microRNA biogenesis via modulation of Lin28B protein. Biochim Biophys Acta 1839:374–386
129. Wang L-X, Wang J, Qu T-T, Zhang Y, Shen Y-F (2014) Reversible acetylation of Lin28 mediated by PCAF and SIRT1. Biochim Biophys Acta 1843:1188–1195

130. Li S, Wang L, Fu B, Berman MA, Diallo A, Dorf ME (2014) TRIM65 regulates microRNA activity by ubiquitination of TNRC6. Proc Natl Acad Sci U S A 111:6970–6975
131. Cheever A, Ceman S (2009) Phosphorylation of FMRP inhibits association with Dicer. RNA 15:362–366
132. Muddashetty RS, Nalavadi VC, Gross C, Yao X, Xing L, Laur O, Warren ST, Bassell GJ (2011) Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. Mol Cell 42:673–688
133. Jee D, Lai EC (2014) Alteration of miRNA activity via context-specific modifications of Argonaute proteins. Trends Cell Biol 24:546–553
134. Wu C, So J, Davis-Dusenbery BN, Qi HH, Bloch DB, Shi Y, Lagna G, Hata A (2011) Hypoxia potentiates microRNA-mediated gene silencing through posttranslational modification of Argonaute2. Mol Cell Biol 31:4760–4774
135. Zeng Y, Sankala H, Zhang X, Graves PR (2008) Phosphorylation of Argonaute 2 at serine-387 facilitates its localization to processing bodies. Biochem J 413:429–436
136. Qi HH, Ongusaha PP, Myllyharju J, Cheng D, Pakkanen O, Shi Y, Lee SW, Peng J, Shi Y (2008) Prolyl 4-hydroxylation regulates Argonaute 2 stability. Nature 455:421–424
137. Narayanan U, Nalavadi V, Nakamoto M, Thomas G, Ceman S, Bassell GJ, Warren ST (2008) S6K1 phosphorylates and regulates fragile X mental retardation protein (FMRP) with the neuronal protein synthesis-dependent mammalian target of rapamycin (mTOR) signaling cascade. J Biol Chem 283:18478–18482
138. Niere F, Wilkerson JR, Huber KM (2012) Evidence for a fragile X mental retardation protein-mediated translational switch in metabotropic glutamate receptor-triggered Arc translation and long-term depression. J Neurosci 32:5924–5936
139. Li Y, Tang W, Zhang L-R, Zhang C-Y (2014) FMRP regulates miR196a-mediated repression of HOXB8 via interaction with the AGO2 MID domain. Mol Biosyst 10:1757–1764
140. Zhang Y, Lu Z, Ku L, Chen Y, Wang H, Feng Y (2003) Tyrosine phosphorylation of QKI mediates developmental signals to regulate mRNA metabolism. EMBO J 22:1801–1810
141. Feng Y, Bankston A (2009) The star family member QKI and cell signaling. Adv Exp Med Biol 693:25–36
142. Abdelmohsen K, Srikantan S, Yang X, Lal A, Kim HH, Kuwano Y, Galban S, Becker KG, Kamara D, de Cabo R et al (2009) Ubiquitin-mediated proteolysis of HuR by heat shock. EMBO J 28:1271–1282
143. Zhou H-L, Geng C, Luo G, Lou H (2013) The p97-UBXD8 complex destabilizes mRNA by promoting release of ubiquitinated HuR from mRNP. Genes Dev 27:1046–1058
144. Li H, Park S, Kilburn B, Jelinek MA, Henschen-Edman A, Aswad DW, Stallcup MR, Laird-Offringa IA (2002) Lipopolysaccharide-induced methylation of HuR, an mRNA-stabilizing protein, by CARM1. Coactivator-associated arginine methyltransferase. J Biol Chem 277:44623–44630
145. Pang L, Tian H, Chang N, Yi J, Xue L, Jiang B, Gorospe M, Zhang X, Wang W (2013) Loss of CARM1 is linked to reduced HuR function in replicative senescence. BMC Mol Biol 14:15
146. Miyagi M, Sakaguchi H, Darrow RM, Yan L, West KA, Aulak KS, Stuehr DJ, Hollyfield JG, Organisciak DT, Crabb JW (2002) Evidence that light modulates protein nitration in rat retina. Mol Cell Proteomics 1:293–303
147. Yogesha SD, Mayfield JE, Zhang Y (2013) Cross-talk of phosphorylation and prolyl isomerization of the C-terminal domain of RNA Polymerase II. Molecules 19:1481–1511
148. Stackpole EE, Akins MR, Fallon JR (2014) N-myristoylation regulates the axonal distribution of the fragile X-related protein FXR2P. Mol Cell Neurosci 62:42–50
149. Leung AKL, Vyas S, Rood JE, Bhutkar A, Sharp PA, Chang P (2011) Poly (ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. Mol Cell 42:489–499

Index

G.W. Yeo (ed.), *RNA Processing*, Advances in Experimental Medicine and Biology 907, DOI 10.1007/978-3-319-29073-7