Resolving RNA complexity to decipher regulatory rules governing biological networks

Donny D. Licatalosi and **Robert B. Darnell**
Howard Hughes Medical Institute, Laboratory of Molecular Neuro-Oncology, The Rockefeller University, 1230 York Ave, NY, NY 10021 USA

Preface

In recent years, views of eukaryotic gene expression have been transformed by the finding that enormous diversity can be generated at the RNA level. Advances in technologies for characterizing RNA populations are revealing increasingly complete descriptions of RNA regulation and complexity—for example through alternative splicing, alternative polyadenylation, and RNA editing. New biochemical strategies to map protein-RNA interactions *in vivo* are yielding transcriptome-wide insights into mechanisms of RNA processing. These advances, combined with bioinformatics and genetic validation, are leading to the generation of functional RNA maps that reveal rules underlying RNA regulation and networks of biologically coherent transcripts. Together, these are providing new insights into molecular cell biology and disease.

Introduction

Gene expression is finely regulated to ensure that the correct complement of RNA and proteins is present in the correct cell at the correct time. Owing to its diversity—in sequence and structure—RNA plays critical roles in cell biology, and is regulated by numerous proteins that modulate its content and spatial-temporal expression. Methodological advances, including bioinformatic, microarray-based, biochemical and deep sequencing studies, are producing new insights into the role that regulation of RNA complexity—the sum of the unique isoforms of RNA in a cell, including mRNA variants, non-coding RNAs and microRNAs (miRNAs)—plays in generating organismal complexity from a relatively small number of genes. Here we review this progress, focusing on mRNAs and the ways in which the technological advances are beginning to revolutionize our ability to understand the mechanisms and consequences of mRNA diversification.

The recognition of RNA regulation as a central point in gene expression and the generation of phenotypic complexity[1] began with new methodologies and biological insights developed in the 1970s–1980s. Nascent transcripts were found to be generated as long heterogeneous nuclear RNAs (hnRNAs)[2,3] (now termed pre-mRNA) that serve as precursors for smaller 5′ capped and 3′ polyadenylated mRNAs that are then exported to the cytoplasm. Insights into the mechanism by which pre-mRNA is processed to mature mRNA resulted from methodologic advances - including S1 nuclease mapping[4] and electron microscopy to visualize R-loops of adenovirus mRNA:DNA hybrids[5,6] - that enabled nucleotide-level examination of the precursor-product relationship of adenoviral transcripts. These efforts revealed that adenoviral mRNA has "an amazing sequence arrangement"[6] such that

Correspondence: Robert B. Darnell, M.D., Ph.D., Laboratory of Molecular Neuro-Oncology, The Rockefeller University, Box 226, 1230 York Avenue, New York, NY 10021, Phone: (212) 327-7460, Fax: (212) 327-7109, darnelr@rockefeller.edu.

6) **Weblinks**
www.rockefeller.edu/labheads/darnellr/

processing of pre-mRNA to mature mRNA involves the intra-molecular joining (splicing) of expressed sequences (exons) that are separated by non-coding intervening sequences (introns)[7] in the primary transcript (Figure 1). This was quickly recognized as a general feature of eukaryotic RNA processing[8,9].

The discovery of splicing led to the realization that RNA has the potential to be more complex than DNA[7,10]. This potential was demonstrated by the finding, first in adenovirus[11] and subsequently in eukaryotic cells during cell differentiation[12] and in different tissues[13], that alternative mRNA products could be generated from a single pre-mRNA precursor in a regulated manner. In this way regulation of alternative splicing and polyadenylation enables a single mammalian gene to encode multiple mRNAs that possess distinct coding and regulatory sequences.

A more recent epoch in understanding RNA complexity was ushered in with the ability to sequence complete genomes, and the concomitant realization that humans and worms have roughly the same number of protein coding genes (and, more recently, that human and chimpanzee genomic coding regions are 99.7% identical)[14]. These observations, together with the development of the RNA World hypothesis[15, 16], led to a new concept that is explored in this review. This concept is that biological complexity—the variation in cell type and function—has RNA complexity at its core. In this view, it is the intricate unfolding of the genetic information in DNA into diverse RNA species - mediated by RNA-protein interactions - that leads to biological variation not evident from analysis of DNA sequence alone.

The known roles of RNA in the cell have expanded from it being a machine and template for protein synthesis to a regulatory hub for post-transcriptional control with emerging, and still incompletely understoood roles as a *trans*-acting factor that is capable of regulating expression of genetic information. For example, miRNAs[17], piRNAs[18] and long non-coding RNAs[19,20] act to direct different RNA binding proteins (RNABPs) to their regulatory targets in order to suppress translation[21], provide protection from transposable elements[18], and mediate epigenetic changes[1,22,23], respectively. Adding to its versatility, RNA transcripts are diversified from the point of transcription onwards through the action of a plethora of mechanisms, including alternative transcription initiation[24–26], alternative splicing[27–29], alternative polyadenylation[30], RNA editing[31], and post-transcriptional modification (pseudouridylation[32], methylation[33], and non-canonical polyadenylation and RNA terminal polyuridylation[34,35]). Once generated, mature RNA isoforms are subject to many levels of regulation that include the regulation of translation by miRNAs[21] and regulatory factors[36], the use of alternative translational start sites[37], RNA localization[38], and mRNA stability and turnover[39,40].

RNA regulation is achieved through the concerted action of multiple RNABPs[41] that bind to 'core' and 'auxiliary' elements, which are required for and modulate pre-mRNA processing events, respectively (Box 1). Core splicing elements demarcate exons and the sequences required for their splicing, and auxiliary splicing elements, which are located in introns and/ or exons, bind factors that enhance or inhibit splicing. Similarly, mRNA 3′ end maturation also depends on the presence of core and auxiliary elements that define the site of transcript cleavage and polyadenylation[42,43]. The identification of alternative polyadenylation sites in the majority of human genes and evidence for tissue-specific biases in alternative polyadenylation[8,44–46], suggests that regulation of alternative polyadenylation through auxiliary control might be a common mechanism to diversify the transcriptome.

BOX 1

Alternative splicing and polyadenylation

Core elements necessary for pre-mRNA splicing include the 5′ and 3′ splice sites (SS), a branch point sequence (BP) upstream of the 3′SS, and a polypyrimidine-rich tract (PPT) between the BP and the 3′ SS. All of these elements are bound by components of the spliceosome, which is a dynamic macromolecular complex that consists of snRNAs and ~170 proteins[29]. Auxiliary sequences are variable in number and location - they can be located in exons and in the flanking intronic sequences - and are bound by factors that generally function to either enhance or inhibit basal splicing activity. The combinatorial actions of both core and auxiliary splicing factors participate in the regulation of alternative splicing. For example, the SR proteins comprise a family of auxiliary RNABPs that bind to splicing enhancer elements to facilitate exon identification and promote splicing (although like most RNABPs they are also able to serve other functions in the cell). In contrast, the binding of auxiliary hnRNP proteins to splicing silencer elements has a negative effect on exon inclusion; in many cases they antagonize the "pro-splicing" activity of SR proteins. Interestingly, the levels of some core snRNPs vary between tissues[128], and such variations might contribute to splicing regulation[41]. Core elements necessary for maturation of the 3′ end of an mRNA include a poly(A) signal (an adenylate-rich hexameric sequence, most often AAUAAA) and a U/GU-rich sequence, which are positioned upstream and downstream of the poly(A) site respectively. These elements direct the endonucleolytic cleavage and polyadenylation of the transcript. Although a number of auxiliary elements that affect the use of poly(A) sites have been identified[43], the extent to which these elements regulate alternative poly(A) site use remains unclear.

Current interest relating to RNA complexity has three main aspects: meeting methodological challenges so that the vast amount of information present in RNA can be collated; analysis of these data sets so that new rules of RNA regulation can be detailed; and application of the new insights, to achieve a basic understanding of cellular control and, ultimately, an understanding of gene dysregulation in human disease. This review will discuss each of these points - methodology, RNA analysis and, more briefly, its biological manifestations – in each case focusing on the control of RNA complexity. Although this review touches on many aspects of RNA function, including links to transcriptional and translational regulation, space does not allow a discussion of these issues, which are discussed in several excellent reviews[19,24,36,41,47–50].

New methods to analyze RNA complexity

Although advances in the 1970s–80s came about through the detailed study of individual RNAs, the focus of recent technological advances is the characterization of whole RNA populations in cellular contexts with nucleotide level resolution. Accordingly, new methods able to simultaneously analyze multiple RNA processing events are culminating in the development of genome-wide RNA maps that pave the way to new biological insights.

Microarrays

Systematic efforts to identify RNA variants began with microarray technologies. A variety of different arrays have been used to elucidate RNA complexity. In particular, probesets for alternative exons identified from genome sequencing efforts have been used to analyze splice variants. The first use of these exon-junction microarrays to interrogate RNA populations from different tissues led to the recognition that a large number (at the time, the

estimate was ~75%, but see below) of human multi-exon genes are alternatively spliced[51]. Similar exon-junction arrays have been used to identify tissue-restricted patterns of alternative mRNA expression and provide insights into their regulation by specific RNABPs[52,53].

The ability of microarrays to provide valuable data on alternative RNA processing has led to their productive use as tools to assess mRNA diversity in different biologic contexts. Nonetheless, microarray studies have been limited by several factors. Two of these factors - the incomplete nature of gene annotations and limitations on microarray density - continually improve over time, but others, such as the need to predefine targets (such as alternative exons), preclude the identification of novel alternative mRNA isoforms. One effort to address this latter issue has been the development of microarrays that can be used to interrogate "complete" sets of transcribed exons[54]. Although these arrays do not monitor specific splice junctions, they have the advantage of expanded transcriptome coverage, which provides more reliable estimates of RNA abundance, and they are able to detect changes in the usage of individual exons (alternatively spliced isoforms) as well as variants derived from differential transcription regulation or alternative polyadenylation. More complete "genome tiling" arrays have been developed for yeast, Drosophila, and some human chromosomes[55,56]; these arrays circumvent the need for prior knowledge of the transcriptome. Analyses using tiling arrays reported that the vast majority of the human genome is transcribed[57], although the biological relevance of these findings remains uncertain[56,58]. A final limitation of microarrays is they dependent on nucleic acid hybridization; researchers need to consider signal-to-noise ratios that can vary owing to the differences in base composition and annealing properties between individual probes. These limitations are being addressed with a new technology, direct high-throughput sequencing.

High throughput sequencing

RNA-Seq (or next-generation RNA sequencing) (Box 2) takes advantage of the power of new single molecule sequencing methods[59,60] that are currently able to produce billions of nucleotides of sequence in a matter of days for several thousand dollars. The power of RNA-Seq to assess mRNA complexity was highlighted in 2008 by the Blencowe[61] and Burge[45] laboratories, who provided complete RNA profiles and analysis of alternative splicing and polyadenylation variants in different tissues that easily rivaled those that could be obtained using microarrays. The ability of RNA-Seq to detect previously uncharacterized mRNA isoforms and new classes of non-coding RNAs[62] illustrate the utility of this rapidly evolving technology, which is assuming an increasingly dominant role in RNA analyses. In addition, high throughput sequencing can be coupled with hybridization strategies to enrich specific RNA populations prior to sequencing. This balances constraints of hybridization technologies (as with microarrays) with the advantages of high throughput sequencing experiments, and has been effectively used to study RNA variants generated by RNA editing[63,64].

BOX 2

RNA-Seq

The term RNA-Seq applies to any of several different high throughput (next-generation) sequencing methods to obtain transcriptome-wide RNA profiles[59]. Typically, RNA from two samples that are to be compared is sheared, converted to cDNA, and sequenced. This can yield, for example, up to 25 million sequence reads that are ~35 nt in length. Although there can be sequencing bias at any particular position in the genome - for example, depending on GC-content and/or the propensity of that sequence to be amplified by PCR - such errors will be the same across different samples. Therefore,

---

differences between samples can be quantitated at the resolution of individual splice variants[45] or even edited RNA nucleotides[63]. Other applications of RNA-Seq, using different sequencing strategies, include looking at pools of RNA that are being translated by sequencing RNA bound to ribosomes[159] and single cell RNA analysis[162]. Currently 2.5×10^7 sequence reads can detect 2.5×10^5 different transcripts. This means that abundant transcripts are represented by many reads and rare transcripts by only a few; the sensitivity of this technique is likely to improve over time.

Bioinformatics

Bioinformatics has emerged as a powerful compliment to current efforts to analyze the complexity of cell-specific RNA signatures. Sequence-based bioinformatic approaches have long been applied to the study of pre-mRNA processing and have revealed consensus sequences that define the 5′ splice site[65], the poly(A) signal that is necessary for 3′ end maturation and termination of transcription[66,67], and atypical consensus sequences that define an entire alternative means for regulating splicing[68]. Current bioinformatic efforts are aided by, and are also dependent on, improvements in the number and depth of sequences available from EST and cDNA libraries, microarray datasets and whole genome sequencing. Therefore bioinformatics is likely to become more powerful as new technology improves such databases.

Comparison of RNA profiles from different cell types and organisms have helped to determine the frequency of alternative processing and the extent to which it is subject to species or tissue-specific regulation. In addition, analyses of sequences associated with conserved alternative processing events have helped to develop understanding of a number of aspects of alternative processing, including: identifying sequence elements that are potentially associated with the regulation of alternative processing[52,69–72]; investigating the origins of alternative splicing[73]; and defining unexpected features, such as ultraconserved elements that mediate nonsense-mediated decay of transcripts that encode RNABPs[74]. Although not the focus of this review, bioinformatics has also been used in efforts to identify miRNA targets[47] and other regulatory elements in 3′ UTRs[38,39,40]

Methods to study protein-RNA interactions

Bioinformatic, microarray, and high throughput sequencing studies have provided an unprecedented ability to describe RNAs on a genome-wide scale and to suggest which *cis* elements and *trans*-acting factors are associated with their regulation. However, these methods are limited without biochemical methods to identify the direct RNABP-RNA interactions that define that regulation *in vivo*. In general, researchers wish to distinguish between the primary (direct) and secondary (indirect) effects of RNA regulatory factors. For example, in the Fragile-X mental retardation syndrome, the loss of FMRP function is clearly the proximate cause of the disorder. Therefore there is great interest in identifying the RNAs that FMRP regulates in neurons and in distinguishing these direct effects from RNA dysregulation that is owing to secondary or tertiary consequences of FMRP loss[75]. Put another way, any perturbation in a cell is likely to disrupt the RNA profile of that cell, as detected by methods such as microarray or RNA-Seq. Therefore, changes in RNA profiles are expected after perturbation of RNABPs, and such changes cannot be taken as evidence of specific action of an RNABP. Attempts to study mechanisms of RNA regulation in cells depend on distinguishing direct from indirect consequences of cellular manipulations.

Multiple approaches have emerged for the biochemical identification of functional RNABP-RNA interactions *in vivo*. These include immunoprecipitation of RNABPs followed by purification of the co-precipitating RNA and analysis by RT-PCR or microarray

analysis[76,77]. These strategies have proven useful, but they cannot discriminate direct from indirect interactions, nor identify RNA-protein binding sites. Moreover, they are limited by the need to use relatively low stringency conditions to maintain protein-RNA interactions, and such conditions are associated with problems related to signal-to-noise ratio, co-precipitating RNABPs and RNABP-RNA re-association *in vitro*[78–80].

An alternative means of identifying regulatory RNABP-RNA interactions is the CLIP (cross-linking and immunoprecipitation) assay[78,81,82] (Box 3). CLIP applies the observation - first made in the study of tRNA-protein interactions in the 1970s[83] and even earlier for DNA-protein interactions - that UV-irradiation causes covalent cross-linking between RNA-protein complexes that are in tight apposition (that is, within ~Ångstrom distances). UV cross-linking was applied in a cellular context in studies of protein-RNA interactions by van Venrooij[84] and Pederson[47,85,86] in the 1980s, and were then refined by immunoprecipitation of cross-linked hnRNP-RNA complexes by Dreyfuss[87] and colleagues.

BOX 3

CLIP and HITS-CLIP methods

CLIP takes advantage of the ability of UV-irradiation to penetrate intact cells or tissues and induce covalent crosslinks between RNA and proteins that are in direct contact (~1 Ångstrom apart). A flow diagram of the experimental steps is shown in the figure. Once they have been covalently bound, RNA-protein complexes can be purified under harsh conditions, which gives the advantage of being able to separate them from closely bound RNABP-RNABP complexes, reassociated RNAs, and background RNA. After purification, the CLIP method[81,82,92] utilizes proteinase K to remove the RNABP. This is followed by linker ligation and RT-PCR to analyze the RNA sequences. This sequencing analysis can be done using high throughput sequencing methods, in which case it is referred to as "HITS-CLIP" [88,91]. The details of HITS-CLIP are likely to be modified and improved over time. For example, more efficient sequencing and the use of ever-smaller sample sizes are likely to be possible. Current methods and algorithms for analyzing HITS-CLIP data can be found at www.rockefeller.edu/labheads/darnellr/. It should be noted that it remains to be determined whether HITS-CLIP has limitations in terms of efficiency of crosslinking specific subsets of RNA-protein interactions. However, to date, microarray and HITS-CLIP studies have yielded very similar results[53,88], which suggests that crosslinking can be highly efficient across the transcriptome.

a Cross-linking
b RNAse   RNAse
   5'-OH — RBP — 3'-P
c Alkaline phosphatase
   5'-OH — RBP — 3'-OH
d RNA ligase
   5'-OH — RBP
e Polynucleotide kinase
   γ^{32} P — RBP
f SDS-PAGE
   RNA-protein
   Free RNA
g Proteinase K
h RNA ligase
i RT-PCR
j High throughput sequencing and genome mapping
   CLIP tags
   Alternative exon
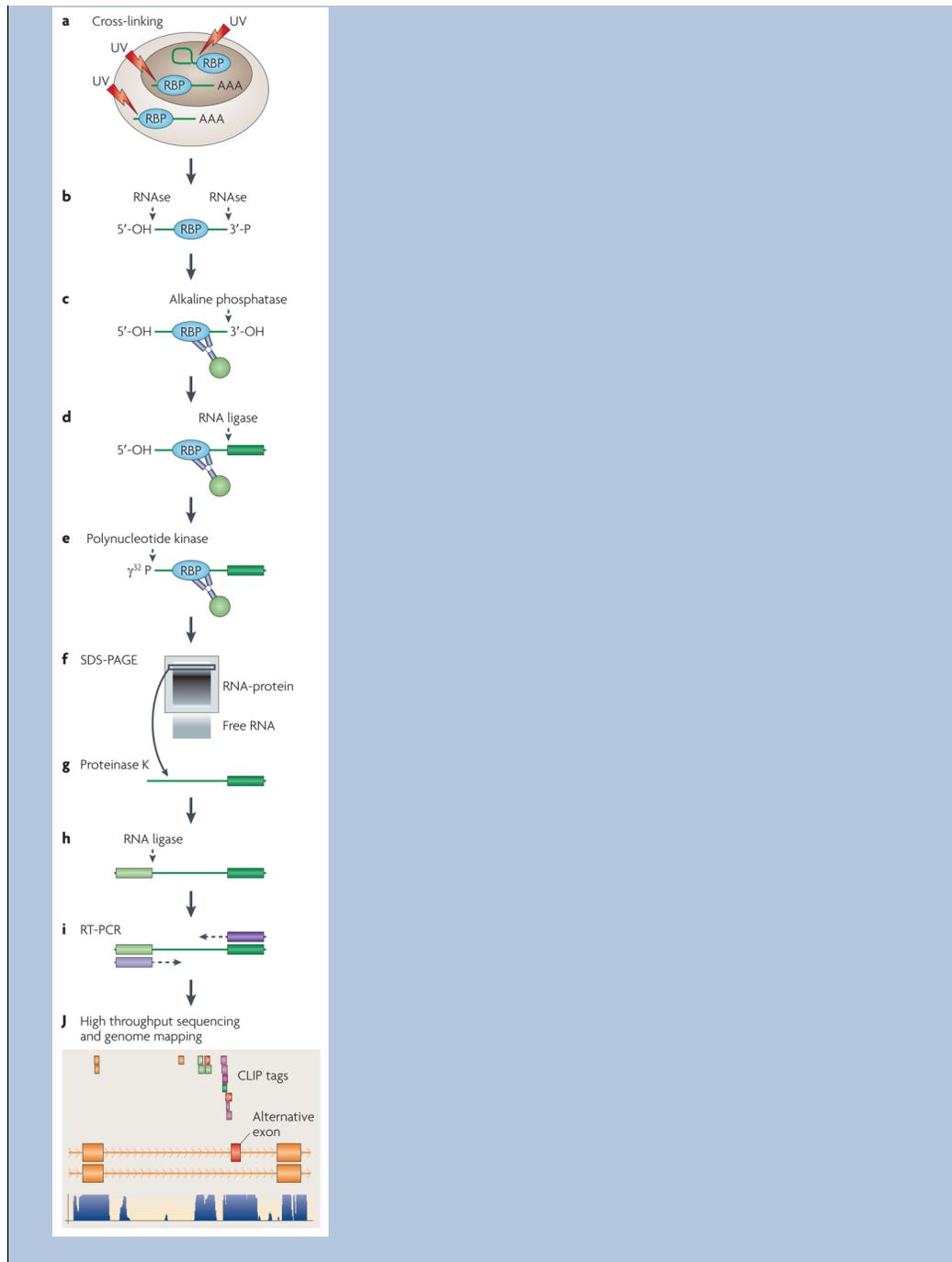
CLIP allows the purification of RNABP-RNA interactions that are occurring in live cells, or even whole tissues such as the brain, to be covalently "locked" in place and rigorously purified. This yields a population of RNA sequences that are directly bound by the RNABP of interest. Sequencing of this population provides a means of identifying the bound RNA, and importantly, the position of protein binding. The CLIP method (and emerging methodological improvements to this method [81,82,88–91]) established that small crosslinked RNA fragments could be amplified by RT-PCR amplification after partial RNase and proteinase K digestion. This approach - initially using conventional strategies[92] and more recently using high throughput sequencing (HITS-CLIP) [88] - reveals the RNA "sequence

footprints" that are bound by RNABPs and provides a powerful way to study RNA-protein interactions in living tissues on a transcriptome-wide level[1,41,59]. So far, HITS-CLIP has been used to generate high resolution genome-wide assessments of RNABP-RNA interactions in mouse brain[88], stem cells[90], and tissue-culture cells[93], and to deconvolute Argonaute-miRNA-mRNA ternary interactions in the mouse brain[91] (discussed further below).

Generating biologic complexity through RNA regulation

As new methods have improved the ability to assess mRNA complexity, estimates of the extent to which alternative RNA isoforms contribute to functional diversity have increased. Recent efforts to characterize the mRNA signature of different human tissues using RNA-Seq have revealed that nearly all multi-exon human genes (comprising >90% of all genes) generate alternative mRNA isoforms, and most do so in a tissue-specific manner[45,61]. These alternative isoforms include variants that arise from alternative transcription initiation and from all known forms of alternative pre-mRNA processing. In addition, high throughput sequencing combined with target enrichment has been used to assess the diversity generated by over 36,000 sites at which RNA editing occurs[63]. All of these means of modifying RNA transcripts generate complexity of both protein coding mRNAs and ncRNAs; we focus here on RNA as the regulated substrate (as opposed to DNA as the substrate, which is reviewed elsewhere[24,26,41]), and note that, to date, most experimental validation has been done with protein coding mRNAs.

Alternative splicing

Alternative splicing is one of the major ways in which RNA diversity is generated. Comparative analysis of splicing variants are yielding insights into its biological consequences[76,94–97]. Interestingly, RNA-Seq-based characterization of tissue transcriptomes, together with microarray analyses[52,54,69] and comparative bioinformatic studies[71,98], identified the mammalian brain as the tissue expressing the greatest number of alternative mRNA isoforms. This is likely to be related to the fact that this tissue is populated by thousands of highly specialized unique cell types that undergo dynamic changes. In the nervous system, alternative splicing has many important roles including controlling the spatial and temporal expression of isoforms necessary for neurodevelopment and modification of synaptic strength[95,99].

Important general issues regarding the complexity of alternative splicing are highlighted by contrasting studies of Dscam (Down Syndrome cell adhesion molecule) and neurexin splicing in the nervous system. In Drosophila, *Dscam* - which is believed to be crucial for proper neural circuit formation - encodes many thousands of neuron-specific RNA variants that are produced by alternative splicing[100]. Despite the great complexity of RNA products, and the recognition that RNABPs act to restrict *Dscam* exon usage[101], it is believed that the choice of RNA variants produced in any one neuron are largely stochastic, and the resulting biologic complexity is proportionately low. Each RNA variant encodes a cell surface axonal molecule that is randomly generated to be different from that on neighboring axons, thereby yielding a unitary outcome - that is, the inability to fasciculate with neighbors[100].

Regulated, rather than stochastic, production of alternative RNA variants, has the potential to generate a great diversity of biological function. Alternative splicing of *neurexin* pre-mRNA in mammalian brain provides an interesting example. Nearly 3000 unique *neurexin* transcripts are derived from the combination of three genes, each of which has two alternate promoters and encodes transcripts with ~10 alternate exons[102]. This set of alternate transcripts encodes variants that give rise to alternative neurexin protein isoforms, which have different interactions with different neuroligan protein isoforms across the synaptic

cleft. This suggests that a 'splice code' might underlie trans-synaptic cell adhesion[103]. There is evidence that a small number of RNABPs might regulate neurexin (and neuroligan) isoforms[95], which may in turn generate diverse biological outcomes[103]. These observations underscore the more general point that alternative splicing plays a major role in biological complexity[28].

Alternative polyadenylation

Although it is clear that alternative processing of pre-mRNA can confer different structural and functional properties to proteins[76,104], additional functional roles for alternative processing in the regulation of gene expression have also emerged. Consistent with EST-based bioinformatic studies[46], RNA-Seq analysis identified tissue-specific biases in the regulation of tandem polyadenylation sites (Figure 2A, Box 1). Unlike the alternative poly(A) site regulation that is coupled to inclusion of an alternative 3′ terminal exon, alternative polyadenylation at tandem poly(A) sites can yield transcripts with identical protein-coding sequences but with different 3′ UTR sequences. This provides the potential for differential regulation of mRNA expression by RNABPs and/or miRNAs (Figure 2A). Exon microarray and RNA expression studies have indicated that such regulation might have important biologic consequences. Proliferating cells—T lymphocytes[105] and tumor cells[106]—harbor shortened 3′ UTRs. By contrast, the brain - a non-proliferative tissue - appears to regulate polyadenylation so that transcripts harbor, on average, longer 3′ UTRs[45,88]. These studies suggest that these differing cell types regulate polyadenylation in opposite ways to allow RNA to escape from, or be subjected to, different levels of regulation. There are likely to be multiple mechanisms of regulation, including miRNA-mediated regulation of translation[105,106], RNA localization[38] and stability[39,40].

Alternative splicing coupled to NMD

A recently recognized example in which alternative processing is coupled to post-transcriptional control is that of alternative splicing events that result in the introduction of a premature termination codon (PTC), which targets mRNA for degradation by nonsense-mediated mRNA decay (NMD)[24,48,107] (Figure 2B). Although EST-based bioinformatic studies[108,109] had suggested that that alternative splicing coupled to NMD (AS-NMD) is a widely used mechanism for controlling RNA abundance, how widespread it is remains unclear. However, AS-NMD has been shown to regulate the expression of many splicing regulatory factors (some in an auto-regulatory manner), including the SR proteins[74], hnRNP proteins[110–112], and core spliceosomal proteins[109]. Interestingly, some of the exons whose splicing or skipping results in a PTC are associated with ultraconserved elements[74,113]. This suggests that AS-NMD might be an evolutionarily ancient mechanism that is used to establish the correct balance of nuclear RNABPs that is necessary to generate cell-type and developmental-stage specific mRNA profiles[110,111].

Regulating RNA complexity

The dependence of pre-mRNA processing events on multiple RNABP-RNA interactions provides multiple steps at which processing can be regulated. Core elements might be directly involved in the regulation of exon usage, through regulation of core factor stoichiometry[29,114]. Although both SR proteins and hnRNP proteins are widely expressed, changes in their stoichiometry can mediate tissue-specific differences in alternatively splicing[115–117]. Additionally, the activity of RNABPs can be regulated by post-translational mechanisms, including phosphorylation and subcellular sequestration in response to cellular or metabolic stress[41,118]. Such mechanisms can convert a general splicing repressor to a sequence-specific splicing activator[119]. Therefore it is not sufficient to rely solely on correlative expression data to build models of RNA regulation *of vivo*.

An additional layer of pre-mRNA regulation is imparted by tissue-specific RNABPs. Multiple examples of highly related factors with non-overlapping patterns of expression have been described, including the Nova proteins (Nova1 and Nova2)[78], the polypyrimidine tract binding proteins[120–122], Elavl (Hu) proteins[123], Fox proteins[72,90], and the CELF and MBNL proteins[97,124]. Although many homologous tissue-restricted factors show high levels of conservation, multiple mechanisms provide each homologue with a unique pattern of expression, which suggests the homologues have distinct functional roles. For example, cross-regulation at the RNA level ensures that Ptbp1 and Ptbp2 have mutually exclusive expression patterns in mouse and human cells, and this is believed to be critical for the regulation of neuronal differentiation[110,111,122]. In general, it is anticipated that the relative amounts of different positive-and negative-acting RNABPs might define a "cellular RNA processing code" that dictates the pattern of processing for each pre-mRNA, so that pre-mRNAs with the same set of regulatory elements can be regulated in a coordinate manner[96,125–127]. As detailed below, the application of new methodologies are advancing these concepts in expected and unexpected ways, and are revealing details of the mechanisms - including *cis* and *trans* acting codes - that underlie the establishment and regulation of cell-specific RNA profiles.

Genome-wide analysis of protein-RNA interactions

Changes in the expression of numerous RNA-regulatory proteins are coincident with changes in tissue and developmental mRNA profiles[97,128]. A challenge for the future is to understand how the expression and activity of these regulatory factors are regulated, and how multiple factors in combination control the fate of transcribed RNA. Computational analyses have shown that alternative processing events are associated with highly conserved sequences and have identified elements that are enriched near regulated processing sites and are therefore likely to be functionally important for protein binding and regulation[45,52,69,71,72,88,95,97,129–132]. However, only some of the enriched elements correspond to sequences that have been demonstrated to be bound by specific RNABPs, and in most cases *in vivo* studies have yet not been performed to test the functional significance of suspected RNABP-RNA interactions. Interestingly, highly conserved intronic sequences that are associated with alternative splicing events are large enough to accommodate many RNABP-RNA interactions, which is consistent with the idea of combinatorial control involving multiple RNABPs[52,70,71].

Recently, the complexity of RNABP action has begun to be addressed by combining genetic models with high throughput biochemical, bioinformatic, and RNA profiling methods. Such studies have been facilitated by the development of animals with genetically modified RNABP expression—mouse knockouts[53,88,133–135], transgenic mice[97], morpholino-treated zebrafish embyros[136], or cultured cells in which the expression of specific RNABPs has been knocked down by siRNA[69,72,90,136–139]. The use of high-throughput methods in conjunction with these models is now allowing the identification and functional validation of RNA-protein interactions on a transcriptome-wide scale.

The generation of transcriptome-wide maps of functional RNABP-RNA interactions are providing insights into the rules by which RNA complexity is regulated. For example, these studies have generated compelling evidence that the position of RNABP-RNA interactions within primary transcripts dictates the functional outcome of alternative pre-mRNA processing events (Figure 3). Initial ideas relating to Nova-mediated RNA regulation in mouse brain were provided by detailed studies of two transcripts studied *in vitro* and tissue culture cells[134,140,141]. Subsequently, a combination of studies in Nova knockout mice[142], including exon junction arrays[53], bioinformatics[69,134] and HITS-CLIP[88], expanded these ideas into a general rule. In this work, and in subsequent studies of the Fox1/2 splicing factor with analogous findings[69,72,90], it was shown that binding of RNABPs within an

alternative exon or the flanking upstream intronic sequence is generally associated with exon skipping, whereas binding of RNABPs to the downstream intronic sequence is generally associated with exon inclusion (Figure 3).

The extent to which such position-dependent regulation is a feature of other RNABPs is not currently known, however there is reason to believe that such interactions with target pre-mRNAs may also prove to play general features of RNABP regulation, from bioinformatic and biochemical studies of other RNABPs, including Mbnl, Celf, Ptbp1[69,97] and several hnRNP proteins (A/B, L, LL, F and H)[41,138]. The application of genetic systems and high throughput approaches to identify transcriptome-wide interactions and assess their functional significance will provide a greater understanding of the mechanisms by which RNABPs act, in isolation and combinatorially, to regulate gene expression.

Mapping functional transcriptome-wide RNABP-RNA interactions in an unbiased manner can reveal unanticipated functions for RNABPs in generating RNA diversity and regulation. For example, HITS-CLIP combined with microarray analysis of wild type and Nova2 knockout mouse brain led to the identification of an unexpected role for Nova2 in regulating alternative polyadenylation in the brain[88]. Such studies illustrate a point previously recognized, albeit not on a transcriptome-wide scale, for SR and hnRNP proteins: that RNABPs cannot be neatly allocated to a single functional category, rather they are multifunctional proteins that participate in many aspects of RNA biochemistry. HITS-CLIP analysis of the SR protein Sfrs1 (previously known as Asf/Sf2) in human embryonic kidney cells revealed an over-representation of binding to mRNAs encoding RNA regulatory proteins, which suggested the possibility of a regulatory loop[93]. Another new aspect of RNABP regulation emerged from CLIP analyses of hnRNP-A1. These studies revealed that hnRNAP-A1 binds to the stem-loop sequences in the miRNA precursor primiR-18a[143] in HeLa cells, and in so doing functions as an auxiliary factor to enhance Drosha-mediated processing to mature miR-18a[144].

Recently, HITS-CLIP was extended to the study of ternary interactions between an RNABP (an Argonaute protein; Ago), RNA and miRNAs[91]. These studies developed a genome-wide map of miRNA binding sites in mouse brain transcripts. Such studies offer a means to resolve the difficulty bioinformatic approaches have had in identifying bona fide miRNA seed sites. In addition, they might also yield new rules of RNA regulation—27% of Ago binding sites appeared to be "orphans" in which no miRNA binding site could be identified. Therefore there might be new rules of miRNA-mRNA interactions that are yet to be elucidated.

RNA networks and biological coherence

Prior to the onset of high throughput methods, a number of observations suggested that some level of biological coherence is established by RNA regulation—the idea that coordinate regulation of RNAs encoding related proteins coordinates biological processes. Observations of biological coherence of RNA regulation during sex determination in Drosophila and iron-response pathways in vertebrate cells in the 1990's were followed by more general hypotheses of functionally coherent networks in yeast, tissue culture cells and mouse brain, as recently discussed[95,127,145]. However, the inability to distinguish direct from indirectly regulated RNAs complicated evaluation of such networks.

Now, the combination of genetic systems, bioinformatics and biochemistry can be used to uncover functional roles and networks of RNABPs by rigorously identifying validated sets of transcripts and the biological functions of the encoded proteins. For example, analysis of RNA from wild type and Nova knockout mouse brains on exon junction microarrays[53] revealed that Nova regulates alternative splicing of a biologically coherent set of transcripts

encoding proteins with synaptic functions[53,88,95]. HITS-CLIP and bioinformatic studies[88] showed that a subset of these transcripts were directly regulated by Nova. This network has been able to predict aspects of Nova physiology in the mouse brain, including roles in inhibitory potentiation in the hippocampus[146] and in motor neuron function[147]. Taken together, these studies provided the first demonstration in mammals of the coordinated activity of an RNABP in a biological network. Similarly, analysis of RNA regulatory defects in mouse knockouts of Sfrs1[135,148], Srp38[133] and the Celf and Mbnl proteins[97,149] are poised to reveal the direct roles that different factors have in generating the specific alternative mRNA isoforms that are necessary for proper tissue development or functionality.

Alternative mRNAs and disease

The importance of methods to probe mRNA complexity and understand its regulation is underscored by the growing list of human diseases that are associated with defects in the expression of alternative mRNA isoforms[27,94]. This list includes diseases that result from mutations that activate cryptic splice sites or disrupt sequences that are necessary for RNA processing, which lead to the alteration of specific protein isoforms or transcript destabilization. Also, there is also a growing list of disorders that show changes in RNABP expression and/or activity owing to mutation, autoimmune targeting or sequestration of RNABPs. Such disorders seems to particularly affect complex tissues, and are exemplified by neurodegenerative disorders. RNABPs that have been linked to neurodegeneration include: FUS and TDP-43, which are mutated in patients with familial ALS[150]; Nova and the Elavl (Hu) proteins, which are targeted by the immune system in paraneoplastic neurodegenerative disorders[78,151]; SMN1, which is mutated in spinal muscular atrophy[27]; IGHMBP2, which is mutated in spinal muscular atrophy and respiratory distress[152]; senataxin, which is mutated in ALS4[153]; and glycyl tRNA synthetase, which is mutated in hereditary motor neuronopathy type V[154]. Moreover, a growing number of neurological disorders are believed to be linked to RNA expansions that sequester RNABPs, as exemplified by the sequestration of MBNL by CUG repeats in myotonic dystrophy[149]. Similarly, the deletion, mutation or inappropriate expression of miRNAs, which leads to mistargeting of the RNABP Ago and to aberrant RNA regulation[17], is important in multiple disorders[27], including neurologic disease, cancer, and autoimmunity. Although we are just beginning to appreciate the role of RNABPs in human disease, methods that allow researchers to overlay RNA sequence profiles and RNABP maps offer a new means of comparing protein-RNA interactions in normal and diseased tissues.

There are also many examples of defects in the expression of alternative mRNA isoforms and RNABPs in disease for which a defined causal relationship has not been shown. For example, microarray and high throughput RT-PCR analyses have detected alternative splicing events associated with different types of cancer and have identified 'splicing signatures' associated with different histologically defined tumor subgroups[155]. It seems likely that the expression of aberrantly spliced transcripts will be found to contribute to tumor biology. Efforts to identify alternative splicing markers associated with disease, combined with bioinformatic analyses, are providing insights into the mechanisms of RNA regulation that, when perturbed, might result in disease. For example, consensus binding sites for the Fox1/2 RNABPs were identified near many alternative exons that were mis-spliced in ovarian and breast cancer[156]. Evidence suggesting that the Fox proteins directly regulate these alternative splicing events include decreased levels of Fox2 in ovarian cancer and the recapitulation of cancer-associated splicing defects by knock-down of Fox2 expression in cultured cells. Such efforts are providing new insights into the extent to which alternative mRNA isoforms correlate with, and in some cases cause, disease and how disruption of RNABPs that have tumor suppressor[156] or proto-oncogene[157] activities might

lead to the aberrant mRNA processing events that are associated with cancer. Considering the many ways in which alternative processing can affect gene expression, the ability to characterize RNA profiles and regulation in disease will likely play a major role in advancing our understanding of the biology of disease and assist in the development of strategies for therapeutic intervention.

Concluding remarks and future directions

Methodological advances in the 20th century led to the realization that RNA complexity and its regulation lies at the core of biologic complexity. In recent years, the advent of high throughput strategies have enabled nucleotide level analyses of RNA regulation and complexity on a genome-wide scale, which have revealed insights into the extent to which mRNA diversification contributes to cell-specific biology, and the mechanisms by which this diversification is achieved. A challenge for the future will be to determine the extent to which different RNA isoforms contribute to biological complexity.

The complimentary methodological approaches that are described in this review that each give powerful but incomplete data about RNA regulation: methods to enumerate RNA variants (microarrays and RNA-Seq) and bioinformatic approaches are correlative, and biochemical crosslinking alone does not yield functional data. Importantly, combining these efforts (Figure 3) offers the opportunity to identify and experimentally investigate different types of RNA regulatory mechanisms. Such studies have revealed that RNABPs regulate biologically coherent RNA networks, and unanticipated mechanisms by which they do so are emerging. The variety of interactions that are evident from genome-wide studies of RNABPs emphasizes that they are multifunctional proteins whose activities depend on affinity constants and local concentrations of proteins and their RNA substrates. Therefore an important consideration for the future will be to consider how RNABPs act in the context of their local environment—nuclear compartments, cytoplasmic P-bodies, stress granules, dendrites, and so on—and the impact that accessibility of RNA targets has on RNABP activity.

Another challenge will be to take individual RNA maps - each based on genetics, bioinformatics and genome-wide biochemistry - and superimpose them to give a more complete picture of how regulation works inside a cell, in which hundreds of RNABPs simultaneously compete to regulate thousands of RNAs. Such pictures will be needed to interpret the dynamics of RNA-protein interactions during biological processes[158]. Analysis of RNA-protein regulatory maps is also likely to yield insight into non-coding RNAs and their roles in coordinating gene regulation. Finally, application of the methods and concepts reviewed here will advance our understanding of other RNA regulatory mechanisms. For example, translational control is beginning to be studied by using high-throughput methods: yeast translation was recently studied by using RNA-Seq[159] to characterize polyribosomal RNA, and mouse genetics coupled to microarray profiles[160,161] was used to profile transcribed mRNAs within individual neuronal subtypes. Combining the methods described in this review with single cell and, ultimately, subcellular analysis will offer the opportunity to understand RNA function in a variety of cellular contexts. Such studies enhance discovery of how RNA regulation impacts tissue complexity and disease by shaping the expression of genetic information.

Online 'at-a-glance' summary

- The limited differences between the genomes of very different species has led to the emerging recognition that biologic diversity is likely to derive in large part from the complexity of RNA. This is evident in the diverse ways that RNA

molecules are generated and processed and the variety of ways in which mRNA expression is regulated.

- Just as new methodologies drove a revolution in our understanding of the role of RNA in biology in the 20th century, new high throughput sequencing, bioinformatics and biochemical methods are now being applied to whole tissues and genetically defined systems to generate new insights into the role of RNA in biologic systems.

- Alternative splicing is one of the best studied mechanisms by which RNA diversity is generated. Regulation of splicing offers a means of generating RNA variants that offer great biologic variability.

- Alternative polyadenylation is emerging as an important means for regulating 3′ UTRs, which in turn offer a variety of means of regulating gene expression, including microRNA-mediated control of translation, RNA localization and turnover.

- The regulation of RNA processing involves a host of regulatory RNABPs that act according to their affinities for different RNA sequences and the local abundance of RNA and protein. Thus a combination of biochemistry and cell biology will be required to fully understand RNA regulation in mammalian cells.

- Genome-wide analysis of RNA-protein interactions can be rigorously approached biochemically using HITS-CLIP; while methods like next-generation sequencing offers a powerful means of quantitating RNA differences to enumerate RNA diversity. Putting the two approaches together, with bioinformatic tools, allows genome-wide functional RNA maps to be generated, and hence new rules of RNA regulation to be discovered.

- An increasing number of human diseases are being found to relate to targeting of RNABPs, either through their mutation, autoimmune targeting or sequestration by RNA expansions. Applying these same genome-wide analyses to human disease tissues offers the possibility to gain new insight into disease pathogenesis and targeted therapeutics.

Acknowledgments

7) Glossary

RNA World hypothesis	an hypothesis based on the finding that RNA can act as both genetic material and an enzyme[16] that suggests that life originated as an RNA-based form
R-loops	Hybrid structures consisting of RNA and DNA in which RNA displaces a DNA strand to hybridize to its complementary DNA sequence. The formation of R-loops was a commonly used method to define the relationship between genes and their RNA products
piRNAs	Piwi-interacting RNAs. Small RNA species that are processed from single-stranded precursor RNAs. They are 25–35 nucleotides in

	length and form complexes with the piwi protein. piRNAs are probably involved in transposon silencing and stem-cell function
RNA editing	The post-transcriptional modification of RNA primary sequence by insertion and/or deletion of specific bases, or the chemical modification of adenosine into inosine (A to I) or cytidine to uridine (C to U)
ultraconserved elements	large sequences in the genome (usually greater than 200nt) that exhibit very high levels of conservation across multiple species
nonsense-mediated decay	(NMD) The process by which mRNAs containing premature termination codons are destroyed to preclude the production of truncated, and potentially deleterious, protein products. NMD is also used in combination with specific alternative splicing events to control the levels of some proteins
seed sequence	short RNA sequences that are bound by and necessary for miRNA-mediated RNA regulation
morpholino	Modified antisense oligonucleotides (morpholinos). Oligomers that normally consist of 25 morpholino subunits, each of which contains one of the four genetic bases linked to a morpholine ring. The oligomers can bind and inactivate selected RNA sequences on the basis of base pairing and steric interference
siRNA	Short interfering RNA (siRNA). RNA molecules that are 21–23 nucleotides long and that are processed from long double-stranded RNAs; they are functional components of the RNAi-induced silencing complex. siRNAs typically target and silence mRNAs by binding perfectly complementary sequences in the mRNA and causing their degradation and/or translation inhibition
exon skipping	exclusion of an exon from the resulting mature mRNA due to direct splicing of the upstream exon to the downstream exon
exon junction microarray	microarray platform containing probesets designed to detect the mRNA sequences (junctions) formed by splicing of one exon to another

References

1. Sharp PA. The centrality of RNA. Cell. 2009; 136:577–580. [PubMed: 19239877]

2. Scherrer K, Latham H, Darnell JE. Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. Proc Natl Acad Sci U S A. 1963; 49:240–248. [PubMed: 13991616]

3. Soeiro R, Vaughan MH, Warner JR, Darnell JEJ. The turnover of nuclear DNA-like RNA in HeLa cells. J Cell Biol. 1968; 39:112–118. [PubMed: 5678443]

4. Berk AJ, Sharp PA. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell. 1977; 12:721–732. [PubMed: 922889]

5. Berget SM, Moore C, Sharp PA. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. 1977; 74:3171–3175. [PubMed: 269380]

6. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell. 1977; 12:1–8. [PubMed: 902310]

7. Gilbert W. Why genes in pieces? Nature. 1978; 271:501. [PubMed: 622185]

8. Breathnach R, Mandel JL, Chambon P. Ovalbumin gene is split in chicken DNA. Nature. 1977; 270:314–319. [PubMed: 593351]

9. Tilghman SM, et al. Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. Proc Natl Acad Sci U S A. 1978; 75:725–729. [PubMed: 273235]

10. Crick F. Split genes and RNA splicing. Science. 1979; 204:264–271. [PubMed: 373120]

11. Darnell JEJ. Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science. 1978; 202:1257–1260. [PubMed: 364651]

12. Rogers J, Wall R. Immunoglobulin RNA rearrangements in B lymphocyte differentiation. Adv Immunol. 1984; 35:39–59. [PubMed: 6431764]

13. Amara SG, Jonas V, Rosenfeld MG, Ong ES, Evans R. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. Nature. 1982; 298:240–244. [PubMed: 6283379]

14. Calarco JA, et al. Global analysis of alternative splicing differences between humans and chimpanzees. Genes Dev. 2007; 21:2963–2975. [PubMed: 17978102]

15. Cech TR. Crawling out of the RNA world. Cell. 2009; 136:599–602. [PubMed: 19239881]

16. Zaug AJ, Cech TR. The intervening sequence RNA of Tetrahymena is an enzyme. Science. 1986; 231:470–475. [PubMed: 3941911]

17. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136:642–655. [PubMed: 19239886]

18. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell. 2009; 136:656–668. [PubMed: 19239887]

19. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009; 10:155–159. [PubMed: 19188922]

20. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009; 136:629–641. [PubMed: 19239885]

21. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet. 2008; 9:102–114. [PubMed: 18197166]

22. Wang X, et al. Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature. 2008; 454:126–130. [PubMed: 18509338]

23. Cam HP, Chen ES, Grewal SI. Transcriptional scaffolds for heterochromatin assembly. Cell. 2009; 136:610–614. [PubMed: 19239883]

24. Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell. 2009; 136:688–700. [PubMed: 19239889]

25. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH. The functional consequences of alternative promoter use in mammalian genomes. Trends Genet. 2008; 24:167–177. [PubMed: 18329129]

26. Perales R, Bentley D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell. 2009; 36:178–191. [PubMed: 19854129]

27. Cooper TA, Wan L, Dreyfuss G. RNA and disease. Cell. 2009; 136:777–793. [PubMed: 19239895]

28. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature. 2002; 418:236–243. [PubMed: 12110900]

29. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009; 136:701–718. [PubMed: 19239890]

30. Lutz CS. Alternative polyadenylation: a twist on mRNA 3′ end formation. ACS Chem Biol. 2008; 3:609–617. [PubMed: 18817380]

31. Bass BL. RNA editing by adenosine deaminases that act on RNA. Annu Rev Biochem. 2002; 71:817–846. [PubMed: 12045112]

32. Terns M, Terns R. Noncoding RNAs of the H/ACA family. Cold Spring Harb Symp Quant Biol. 2006; 71:395–405. [PubMed: 17381322]

33. Rottman FM, Bokar JA, Narayan P, Shambaugh ME, Ludwiczak R. N6-adenosine methylation in mRNA: substrate specificity and enzyme complexity. Biochimie. 1994; 76:1109–1114. [PubMed: 7748945]

34. Guschina E, Benecke BJ. Specific and non-specific mammalian RNA terminal uridylyl transferases. Biochim Biophys Acta. 2008; 1779:281–285. [PubMed: 18067875]

35. Martin G, Keller W. RNA-specific ribonucleotidyl transferases. RNA. 2007; 13:1834–1849. [PubMed: 17872511]

36. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell. 2009; 136:731–745. [PubMed: 19239892]

37. Kochetov AV. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. Bioessays. 2008; 30:683–691. [PubMed: 18536038]

38. Rodriguez AJ, Czaplinski K, Condeelis JS, Singer RH. Mechanisms and cellular roles of local protein synthesis in mammalian cells. Curr Opin Cell Biol. 2008; 20:144–149. [PubMed: 18378131]

39. Houseley J, Tollervey D. The many pathways of RNA degradation. Cell. 2009; 136:763–776. [PubMed: 19239894]

40. Richter JD. Breaking the code of polyadenylation-induced translation. Cell. 2008; 132:335–337. [PubMed: 18267064]

41. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol. 2009

42. Zhao J, Hyman L, Moore C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev. 1999; 63:405–445. [PubMed: 10357856]

43. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. Cell Mol Life Sci. 2008; 65:1099–1122. [PubMed: 18158581]

44. Manley JL, Takagaki Y. The end of the message--another link between yeast and mammals. Science. 1996; 274:1481–1482. [PubMed: 8966619]

45. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008; 456:470–476. [PubMed: 18978772]

46. Zhang H, Lee JY, Tian B. Biased alternative polyadenylation in human tissues. Genome Biol. 2005; 6:R100. [PubMed: 16356263]

47. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–233. [PubMed: 19167326]

48. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature. 2002; 416:499–506. [PubMed: 11932736]

49. Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. Cell. 2009; 136:719–730. [PubMed: 19239891]

50. Kornblihtt AR. Promoter usage and alternative splicing. Curr Opin Cell Biol. 2005; 17:262–268. [PubMed: 15901495]

51. Johnson JM, et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science. 2003; 302:2141–2144. [PubMed: 14684825]

52. Sugnet CW, et al. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. PLoS Comput Biol. 2006; 2:e4. [PubMed: 16424921]

53. Ule J, et al. Nova regulates brain-specific splicing to shape the synapse. Nat Genet. 2005; 37:844–852. [PubMed: 16041372]

54. Clark TA, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 2007; 8:R64. [PubMed: 17456239]

55. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet. 2005; 21:93–102. [PubMed: 15661355]

56. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. Nat Rev Genet. 2007; 8:413–423. [PubMed: 17486121]

57. Kapranov P, et al. Large-scale transcriptional activity in chromosomes 21 and 22. Science. 2002; 296:916–919. [PubMed: 11988577]

58. Seila AC, Core LJ, Lis JT, Sharp PA. Divergent transcription: A new feature of active promoters. Cell Cycle. 2009:8.

59. Blencowe BJ, Ahmad S, Lee LJ. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. Genes Dev. 2009; 23:1379–1386. [PubMed: 19528315]

60. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009; 10:57–63. [PubMed: 19015660]

61. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet. 2008; 40:1413–1415. [PubMed: 18978789]

62. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458:223–227. [PubMed: 19182780]

63. Li JB, et al. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science. 2009; 324:1210–1213. [PubMed: 19478186]

64. Porreca GJ, et al. Multiplex amplification of large sets of human exons. Nat Methods. 2007; 4:931–936. [PubMed: 17934468]

65. Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. Are snRNPs involved in splicing? Nature. 1980; 283:220–224. [PubMed: 7350545]

66. Logan J, Falck-Pedersen E, Darnell JEJ, Shenk T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc Natl Acad Sci U S A. 1987; 84:8306–8310. [PubMed: 3479794]

67. Proudfoot NJ, Brownlee GG. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature. 1976; 263:211–214. [PubMed: 822353]

68. Patel AA, Steitz JA. Splicing double: insights from the second spliceosome. Nat Rev Mol Cell Biol. 2003; 4:960–970. [PubMed: 14685174]

69. Castle JC, et al. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. Nat Genet. 2008; 40:1416–1425. [PubMed: 18978788]

70. Sorek R, Ast G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res. 2003; 13:1631–1637. [PubMed: 12840041]

71. Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A. 2005; 102:2850–2855. [PubMed: 15708978]

72. Zhang C, et al. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox–2. Genes Dev. 2008; 22:2550–2563. [PubMed: 18794351]

73. Ast G. How did alternative splicing evolve? Nat Rev Genet. 2004; 5:773–782. [PubMed: 15510168]

74. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007; 446:926–929. [PubMed: 17361132]

75. Darnell JC, Mostovetsky O, Darnell RB. FMRP RNA targets: identification and validation. Genes Brain Behav. 2005; 4:341–349. [PubMed: 16098133]

76. Moore MJ, Silver PA. Global analysis of mRNA splicing. RNA. 2008; 14:197–203. [PubMed: 18083834]

77. Tenenbaum SA, Carson CC, Lager PJ, Keene JD. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci U S A. 2000; 97:14085–1490. [PubMed: 11121017]

78. Darnell RB. Developing global insight into RNA regulation. Cold Spring Harb Symp Quant Biol. 2006; 71:321–327. [PubMed: 17381312]

79. Kirino Y, Mourelatos Z. Site-specific crosslinking of human microRNPs to RNA targets. RNA. 2008; 14:2254–2259. [PubMed: 18719246]

80. Mili S, Steitz JA. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA. 2004; 10:1692–1694. [PubMed: 15388877]

81. Jensen KB, Darnell RB. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. Methods Mol Biol. 2008; 488:85–98. [PubMed: 18982285]

82. Ule J, Jensen K, Mele A, Darnell RB. CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods. 2005; 37:376–386. [PubMed: 16314267]

83. Schoemaker HJ, Schimmel PR. Photo-induced joining of a transfer RNA with its cognate aminoacyl-transfer RNA synthetase. J Mol Biol. 1974; 84:503–513. [PubMed: 4840999]

84. Wagenmakers AJ, Reinders RJ, van Venrooij WJ. Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. Eur J Biochem. 1980; 112:323–330. [PubMed: 7460925]

85. Economidis IV, Pederson T. Structure of nuclear ribonucleoprotein: heterogeneous nuclear RNA is complexed with a major sextet of proteins in vivo. Proc Natl Acad Sci U S A. 1983; 80:1599–1602. [PubMed: 6572923]

86. Mayrand S, Setyono B, Greenberg JR, Pederson T. Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)+ heterogeneous nuclear RNA in living HeLa cells. J Cell Biol. 1981; 90:380–384. [PubMed: 6169730]

87. Dreyfuss G, Choi YD, Adam SA. Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. Mol Cell Biol. 1984; 4:1104–1114. [PubMed: 6204191]

88. Licatalosi DD, et al. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature. 2008; 456:464–469. [PubMed: 18978773]

89. Wang Z, Tollervey J, Briese M, Turner D, Ule J. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. Methods. 2009; 48:287–293. [PubMed: 19272451]

90. Yeo GW, et al. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol. 2009

91. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009

92. Ule J, et al. CLIP identifies Nova-regulated RNA networks in the brain. Science. 2003; 302:1212–1215. [PubMed: 14615540]

93. Sanford JR, et al. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res. 2009; 19:381–394. [PubMed: 19116412]

94. Licatalosi DD, Darnell RB. Splicing regulation in neurologic disease. Neuron. 2006; 52:93–101. [PubMed: 17015229]

95. Ule J, Darnell RB. RNA binding proteins and the regulation of neuronal synaptic plasticity. Curr Opin Neurobiol. 2006; 16:102–110. [PubMed: 16418001]

96. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet. 2007; 8:749–761. [PubMed: 17726481]

97. Kalsotra A, et al. A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. Proc Natl Acad Sci U S A. 2008; 105:20333–20338. [PubMed: 19075228]

98. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res. 2002; 30:3754–3766. [PubMed: 12202761]

99. Li Q, Lee JA, Black DL. Neuronal regulation of alternative pre-mRNA splicing. Nat Rev Neurosci. 2007; 8:819–831. [PubMed: 17895907]

100. Hattori D, Millard SS, Wojtowicz WM, Zipursky SL. Dscam-mediated cell recognition regulates neural circuit formation. Annu Rev Cell Dev Biol. 2008; 24:597–620. [PubMed: 18837673]

101. Olson S, et al. A regulator of Dscam mutually exclusive splicing fideltity. Nat Struct Mol Biol. 2007; 14:1134–1140. [PubMed: 21188797]

102. Missler M, Sudhof TC. Neurexins: three genes and 1001 products. Trends Genet. 1998; 14:20–26. [PubMed: 9448462]

103. Boucard AA, Chubykin AA, Comoletti D, Taylor P, Sudhof TC. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron. 2005; 48:229–236. [PubMed: 16242404]

104. Birzele F, Csaba G, Zimmer R. Alternative splicing and protein structure evolution. Nucleic Acids Res. 2008; 36:550–558. [PubMed: 18055499]

105. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. Science. 2008; 320:1643–1647. [PubMed: 18566288]

106. Mayr C, Bartel DP. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009; 138:673–684. [PubMed: 19703394]

107. Lejeune F, Maquat LE. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol. 2005; 17:309–315. [PubMed: 15901502]

108. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003; 100:189–192. [PubMed: 12502788]

109. Saltzman AL, et al. Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol Cell Biol. 2008; 28:4320–4330. [PubMed: 18443041]

110. Boutz PL, et al. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev. 2007; 21:1636–1652. [PubMed: 17606642]

111. Makeyev EV, Zhang J, Carrasco MA, Maniatis T. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell. 2007; 27:435–448. [PubMed: 17679093]

112. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell. 2004; 13:91–100. [PubMed: 14731397]

113. Ni JZ, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 2007; 21:708–718. [PubMed: 17369403]

114. Yu Y, et al. Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. Cell. 2008; 135:1224–1236. [PubMed: 19109894]

115. Caceres JF, Stamm S, Helfan DM, Krainer AR. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science. 1994; 265:1706–1709. [PubMed: 8085156]

116. Ge H, Manley JL. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell. 1990; 62:25–34. [PubMed: 2163768]

117. Krainer AR, Conway GC, Kozak D. The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. Cell. 1990; 62:35–42. [PubMed: 2364434]

118. Biamonti G, Caceres JF. Cellular stress and RNA splicing. Trends Biochem Sci. 2009; 34:146–153. [PubMed: 19208481]

119. Feng Y, Chen M, Manley JL. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat Struct Mol Biol. 2008; 15:1040–1048. [PubMed: 18794844]

120. Markovtsov V, et al. Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol. 2000; 20:7463–7479. [PubMed: 11003644]

121. Polydorides AD, Okano HJ, Yang YY, Stefani G, Darnell RB. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97:6350–6355. [PubMed: 10829067]

122. Spellman R, Llorian M, Smith CW. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. Mol Cell. 2007; 27:420–434. [PubMed: 17679092]

123. Okano HJ, Darnell RB. A hierarchy of Hu RNA binding proteins in developing and adult neurons. J Neurosci. 1997; 17:3024–3037. [PubMed: 9096138]

124. Vicente M, et al. Muscleblind isoforms are functionally distinct and regulate alpha-actinin splicing. Differentiation. 2007; 75:427–440. [PubMed: 17309604]

125. Matlin AJ, Clark F, Smith CW. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol. 2005; 6:386–398. [PubMed: 15956978]

126. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA. 2008; 14:802–813. [PubMed: 18369186]

127. Keene JD. RNA regulons: coordination of post-transcriptional events. Nat Rev Genet. 2007; 8:533–543. [PubMed: 17572691]

128. Grosso AR, et al. Tissue-specific splicing factor gene expression signatures. Nucleic Acids Res. 2008; 36:4823–4832. [PubMed: 18653532]

129. Das D, et al. A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. Nucleic Acids Res. 2007; 35:4845–4857. [PubMed: 17626050]

130. Fagnani M, et al. Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biol. 2007; 8:R108. [PubMed: 17565696]

131. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA. 2005; 11:1485–1493. [PubMed: 16131587]

132. Jelen N, Ule J, Zivin M, Darnell RB. Evolution of Nova-dependent splicing regulation in the brain. PLoS Genet. 2007; 3:1838–1847. [PubMed: 17937501]

133. Feng Y, et al. SRp38 regulates alternative splicing and is required for Ca(2+) handling in the embryonic heart. Dev Cell. 2009; 16:528–538. [PubMed: 19386262]

134. Ule J, et al. An RNA map predicting Nova-dependent splicing regulation. Nature. 2006; 444:580–586. [PubMed: 17065982]

135. Xu X, et al. ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. Cell. 2005; 120:59–72. [PubMed: 15652482]

136. Calarco JA, et al. Regulation of vertebrate nervous system-specific alternative splicing and development by an SR-related protein. Cell. 2009 in press.

137. Aznarez I, et al. A systematic analysis of intronic sequences downstream of 5′ splice sites reveals a widespread role for U-rich motifs and TIA1/TIAL1 proteins in alternative splicing regulation. Genome Res. 2008; 18:1247–1258. [PubMed: 18456862]

138. Blanchette M, et al. Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell. 2009; 33:438–449. [PubMed: 19250905]

139. Venables JP, et al. Multiple and specific mRNA processing targets for the major human hnRNP proteins. Mol Cell Biol. 2008; 28:6033–6043. [PubMed: 18644864]

140. Dredge BK, Darnell RB. Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. Molecular & Cellular Biology. 2003; 23:4687–4700. [PubMed: 12808107]

141. Dredge BK, Stefani G, Engelhard CC, Darnell RB. Nova autoregulation reveals dual functions in neuronal splicing. EMBO J. 2005; 24:1608–1620. [PubMed: 15933722]

142. Jensen KB, et al. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron. 2000; 25:359–371. [PubMed: 10719891]

143. Guil S, Caceres JF. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat Struct Mol Biol. 2007; 14:591–596. [PubMed: 17558416]

144. Michlewski G, Guil S, Semple CA, Caceres JF. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol Cell. 2008; 32:383–393. [PubMed: 18995836]

145. Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol. 2008; 6:e255. [PubMed: 18959479]

146. Huang CS, et al. Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell. 2005; 123:105–118. [PubMed: 16213216]

147. Ruggiu M, et al. Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing. Proc Natl Acad Sci U S A. 2009; 106:3513–3518. [PubMed: 19221030]

148. Kanadia RN, Clark VE, Punzo C, Trimarchi JM, Cepko CL. Temporal requirement of the alternative-splicing factor Sfrs1 for the survival of retinal neurons. Development. 2008; 135:3923–3933. [PubMed: 18987029]

149. O'Rourke JR, Swanson MS. Mechanisms of RNA-mediated disease. J Biol Chem. 2009; 284:7419–7423. [PubMed: 18957432]

150. Lagier-Tourenne C, Cleveland DW. Rethinking ALS: the FUS about TDP-43. Cell. 2009; 136:1001–1004. [PubMed: 19303844]

151. Darnell RB, Posner JB. Paraneoplastic syndromes involving the nervous system. N Engl J Med. 2003; 349:1543–1554. [PubMed: 14561798]

152. Grohmann K, et al. Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. Nat Genet. 2001; 29:75–77. [PubMed: 11528396]

153. Chen YZ, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am J Hum Genet. 2004; 74:1128–1135. [PubMed: 15106121]

154. Antonellis A, et al. Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. Am J Hum Genet. 2003; 72:1293–1299. [PubMed: 12690580]

155. Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer. EMBO Rep. 2008; 9:1087–1093. [PubMed: 18846105]

156. Venables JP, et al. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol. 2009

157. Karni R, et al. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol. 2007; 14:185–193. [PubMed: 17310252]

158. Richter JD, Klann E. Making synaptic plasticity and memory last: mechanisms of translational regulation. Genes Dev. 2009; 23:1–11. [PubMed: 19136621]

159. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009; 324:218–223. [PubMed: 19213877]

160. Doyle JP, et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell. 2008; 135:749–762. [PubMed: 19013282]

161. Heiman M, et al. A translational profiling approach for the molecular characterization of CNS cell types. Cell. 2008; 135:738–748. [PubMed: 19013281]

162. Tang F, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009; 6:377–382. [PubMed: 19349980]

Biographies

Donny D. Licatalosi

Dr. Licatalosi received his Ph.D from the Department of Biochemistry and Molecular Genetics at the University of Colorado Health Science Center, Denver, USA, and is currently a postdoctoral associate in Dr. Robert Darnell's lab at the Rockefeller University, New York, USA. His research interests include understanding RNA-based mechanisms that control the expression of genetic information.

Robert B. Darnell

Dr. Darnell received his undergraduate degree in Biology and Chemistry at Columbia University, New York, USA, and his MD and PhD in molecular biology from Washington University School of Medicine, St. Louis, USA. After finishing clinical training as Chief Neurology Resident at Cornell, Ithaca, USA, and Memorial Sloan-Kettering Cancer Center, New York, USA, he joined the Rockefeller University in 2002, where he is the Heilbrunn Cancer Professor, Senior Physician, and an Investigator of the Howard Hughes Medical Institute. He pioneered studies of the paraneoplastic neurologic disorders and his work led to

the discovery of several genes encoding neuron-specific RNA binding proteins, which he is studying in the brain with innovative methods.
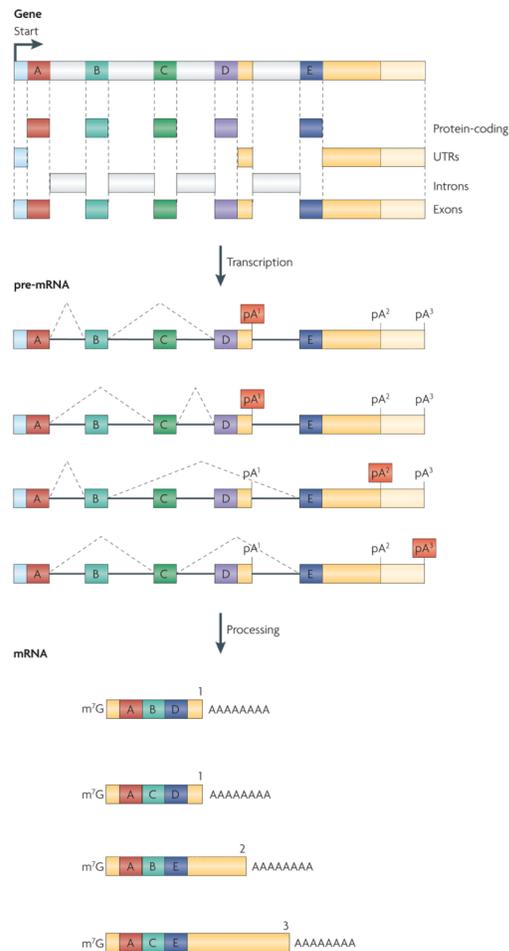
Figure 1. Alternative pre-mRNA processing permits a single gene to encode multiple mRNA isoforms

In this example a single gene generates pre-mRNAs that are alternatively processed to yield mRNA isoforms with different coding and 3′ untranslated regions. Alternative protein-coding regions are established through mutually exclusive splicing of 'B' and 'C' exons and selection of one of two possible 3′ terminal exons ('D' and 'E'). Further mRNA diversification can result from alternative selection of poly(A) sites (pA) in the same 3′ terminal exon (pA2 vs pA3 in the 'E' exon) generating mRNA isoforms with a short or long 3′ UTR. Additional events (not shown) can further diversify the resulting mRNA pool including transcription initiation at an alternative promoter, selection of alternative 3′ or 5′ splice sites (which change exon length), intron retention, and RNA editing.

a Core interactions

b Splicing regulation through combinations of auxiliary factors

c 3′ end maturation
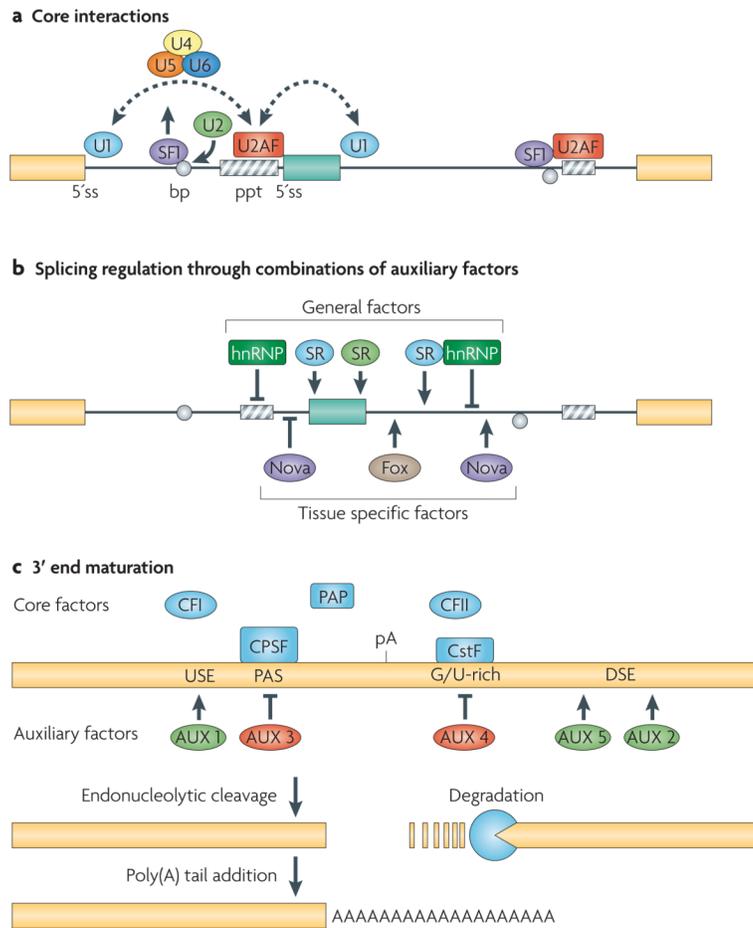
Figure 2. Coupling of RNA processing to alternative RNA regulation

A. Alternative polyadenylation can generate mRNAs with common and isoform-specific 3′ UTR sequences. Changes in 3′ UTR length can alter the repertoire of regulatory elements present in the UTR, such as miRNA target sequences, thus affecting the ability of the transcript to be subject to different forms of post-transcriptional regulation, in this case mRNA degradation. B. Alternative splicing can lead to coding frameshifts, resulting in the introduction of premature translation termination codon (PTC). The presence of a PTC triggers degradation of the mRNA by the NMD pathway, thus regulation of alternative pre-mRNA splicing can be used to control transcript abundance, as evidenced in the physiologic regulation of the RNABPs Ptbp1 and Ptbp2[110–112,122].

a Coupling RNA processing to alternative RNA regulation

b Coupling alternative splicing with nonsense-mediated decay
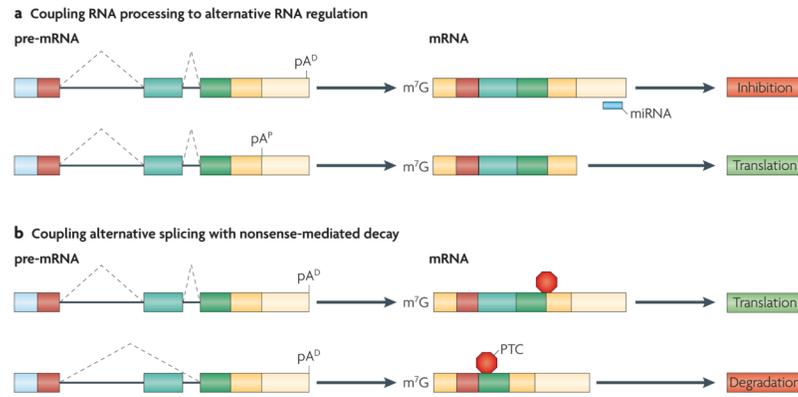
Figure 3. Synergies between methodologies lead to new rules of RNA regulation
Biochemical methods, exemplified by HITS-CLIP, can yield genome-wide footprints of
direct RNA-protein interactions, but lack functional information. In contrast, microarrays or
RNA-Seq are able to correlate differences in RNA profiles between tissues[45] or genetic
systems such as KO vs. WT animals[53], but cannot distinguish direct from indirect targets.
Bioinformatic analysis can also be used to identify sequence features associated with
specific RNA regulatory events[69], but still require biochemical validation of putative
regulatory interactions. Overlaying these approaches can yield a powerful maps of
functional RNA-protein interaction sites for Nova[53,88,134] and Fox1/2[72,90,139].
Two important maps can be derived from combining these approaches, one biologic (bottom
right panel), one mechanistic (bottom left panels). An assessment of the directly regulated
mRNAs can address the extent to which there is a biologically coherence to the set of target
RNAs—for example, the first such assessment of a genome-wide, directly regulated
validated set of targets revealed that Nova regulates RNAs encoding synaptic
functions[53,88,95]. Second, new rules of regulation can be derived from combining
experiments to yield functional maps—for example it became apparent that the position of
binding in a transcript determines the outcome of Nova[134] or Fox1/2[69,72,90] to enhance or
inhibit alternative exon inclusion.
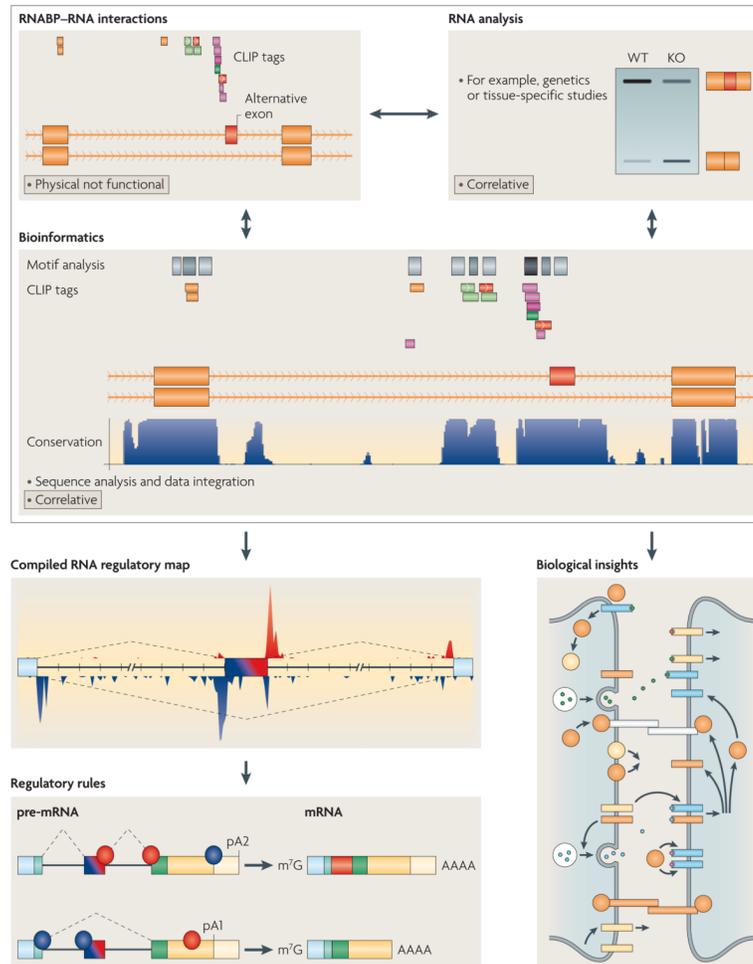
Figure 4. Synergies between methods lead to new rules of RNA regulation
Biochemical methods as exemplified by HITS-CLIP can yield genome-wide footprints of
direct RNA–protein interactions but lack functional information. By contrast, microarrays or
RNA–seq can correlate differences in RNA profiles between tissues[45] or genetic systems,
such as knockout (KO) and wild-type (WT) animals[53], but cannot distinguish direct from
indirect targets. Bioinformatic analysis can also be used to identify sequence features
associated with specific RNA regulatory events[69] but still requires biochemical validation of
putative regulatory interactions. Overlaying these approaches can yield powerful maps of
functional RNA–protein interaction sites for neuro-oncological ventral antigen
(Nova)[53,88,134] and FOX1/2 (REFS [72,90,139]) proteins. Two important maps can be derived
from combining these approaches; one biological (bottom right panel) and one mechanistic
(bottom left panels). An assessment of the directly regulated mRNAs can address the extent
to which there is a biological coherence to the set of target RNAs; for example, the first such
assessment of a genome-wide, directly regulated validated set of targets revealed that Nova
regulates RNAs that encode synaptic functions[53,88,95]. In addition, new rules of regulation
can be derived from combining experiments to yield functional maps; for example, it
became apparent that the position of protein binding in a transcript determines whether
Nova[134] or FOX1/2 (REFS [69,72,90]) binding enhances or inhibits the inclusion of alternative
exons. m^7G, 5′ cap; pA, poly(A) site.