

 APPLICATIONS OF NEXT-GENERATION SEQUENCING

RNA processing and its regulation: global insights into biological networks

Donny D. Licatalosi and Robert B. Darnell

Abstract | 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 the rules underlying RNA regulation and networks of biologically coherent transcripts. Together these are providing new insights into molecular cell biology and disease.

R-loop

A hybrid structure 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 key method to define the relationship between genes and their RNA products.

'RNA World' hypothesis

A hypothesis that life originated as an RNA-based form, based on the finding that RNA can act as both genetic material and an enzyme.

Gene expression is finely regulated to ensure that the correct complement of RNA and proteins is present in the right cell at the correct time. Owing to its diversity — in sequence and structure — RNA has crucial 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 roles that the regulation of RNA complexity — the sum of the unique isoforms of RNA in a cell, including mRNA variants, non-coding RNAs and microRNAs (miRNAs) — has 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 complexity¹ began with new methodologies and biological insights developed in the 1970s and 1980s. Nascent transcripts were found to be generated as long heterogeneous nuclear RNAs (hnRNAs, now termed pre-mRNAs)^{2,3} 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 methodological advances — including S1 nuclease mapping⁴ and electron microscopy to visualize the R-loops

of adenovirus mRNA–DNA hybrids^{5,6} — that allowed the relationship between the precursors and mRNA products of adenoviral transcripts to be examined at the nucleotide level. These efforts revealed that adenoviral mRNA has “an amazing sequence arrangement” (REF. 6), such that the processing of pre-mRNA to mature mRNA involves the intramolecular joining (splicing) of expressed sequences (exons) that are separated by non-coding intervening sequences (introns)⁷ in the primary transcript (FIG. 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 shown by the findings, first made in adenovirus¹¹ and subsequently in eukaryotic cells during cell differentiation¹² and in tissues¹³, 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 allows a single 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 approximately 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

Howard Hughes Medical Institute, Laboratory of Molecular Neuro-Oncology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. Correspondence to R.B.D. e-mail: darnelr@rockefeller.edu doi:10.1038/nrg2673

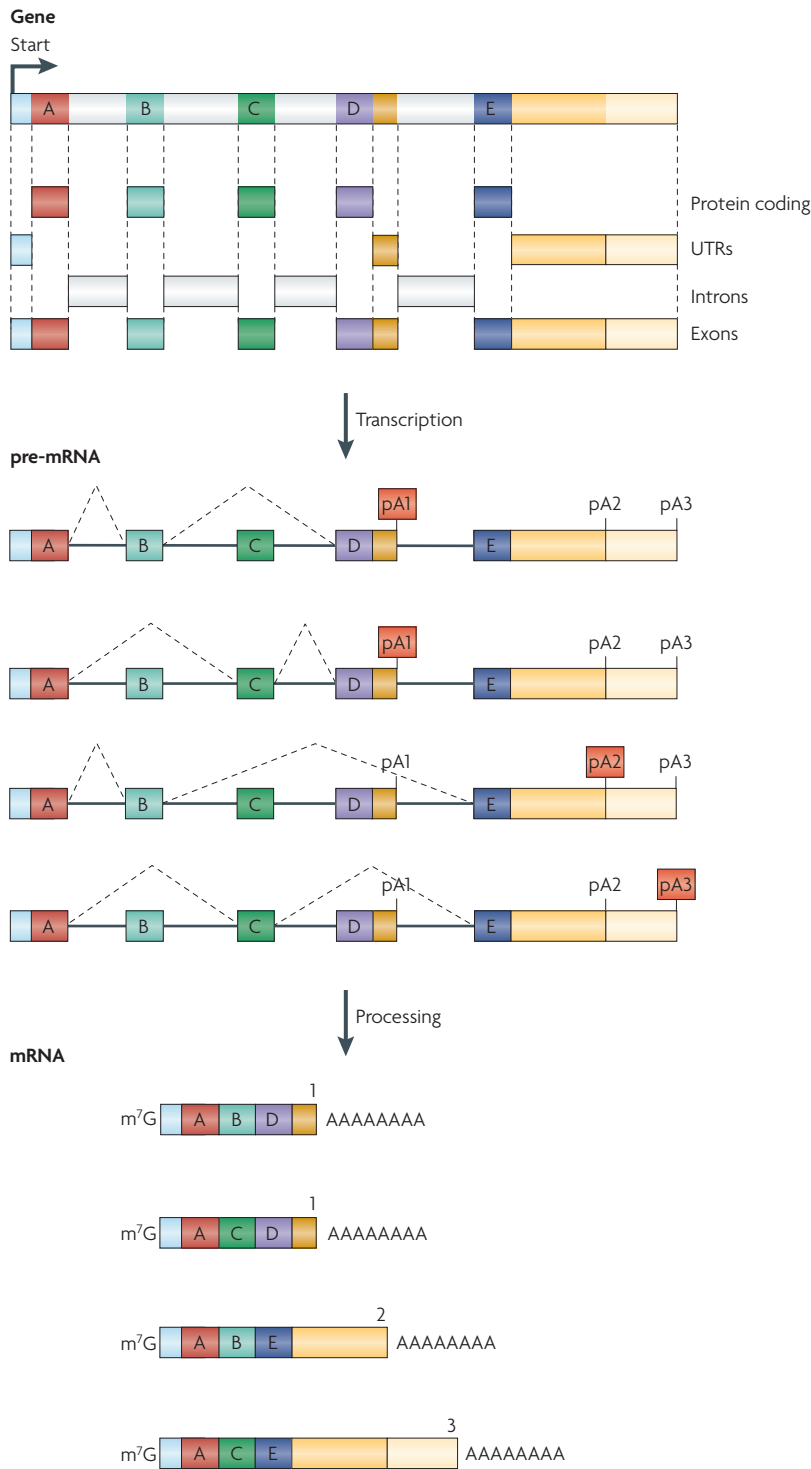


Figure 1 | Alternative pre-mRNA processing allows a single gene to encode multiple mRNA isoforms. In this example a single gene generates pre-mRNAs that are alternatively processed to yield mRNA isoforms with different coding and 3' UTRs. Alternative protein-coding regions are established through mutually exclusive splicing of the '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 versus 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. m⁷G, 5' cap.

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 that is not evident from the analysis of DNA sequence alone.

The known roles of RNA in the cell have expanded from RNA being a machine and template for protein synthesis to it acting as a regulatory hub for post-transcriptional control. There are also emerging and still incompletely understood roles of RNA as a *trans*-acting factor that can regulate the expression of genetic information. For example, miRNAs¹⁷, piwi-interacting RNAs (piRNAs)¹⁸ and long non-coding RNAs^{19,20} direct different RNA-binding proteins (RNABPs) to their regulatory targets to suppress translation²¹, provide protection from transposable elements¹⁸ and mediate epigenetic changes^{1,22,23}, respectively. Adding to the versatility of RNA, transcripts are diversified from the point of transcription onwards through a plethora of mechanisms, including alternative transcription initiation^{24–26}, alternative splicing^{27–29}, alternative polyadenylation³⁰, RNA editing³¹ and post-transcriptional modification (pseudouridylation³², methylation³³ 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 miRNAs²¹ and regulatory factors³⁶, the use of alternative translational start sites³⁷, RNA localization³⁸ and mRNA stability and turnover^{39,40}.

RNA regulation is achieved through the concerted action of multiple RNABPs⁴¹ that bind to 'core' and 'auxiliary' elements, which are required for and modulate pre-mRNA processing events, respectively (FIG. 2). Core splicing elements demarcate exons and the sequences required for their splicing. 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 most human genes and evidence for tissue-specific biases in alternative polyadenylation^{8,44–46} suggests that the regulation of alternative polyadenylation through auxiliary control might be a common mechanism to diversify the transcriptome. 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 deregulation 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 can be found in several excellent reviews^{19,24,36,41,47–50}.

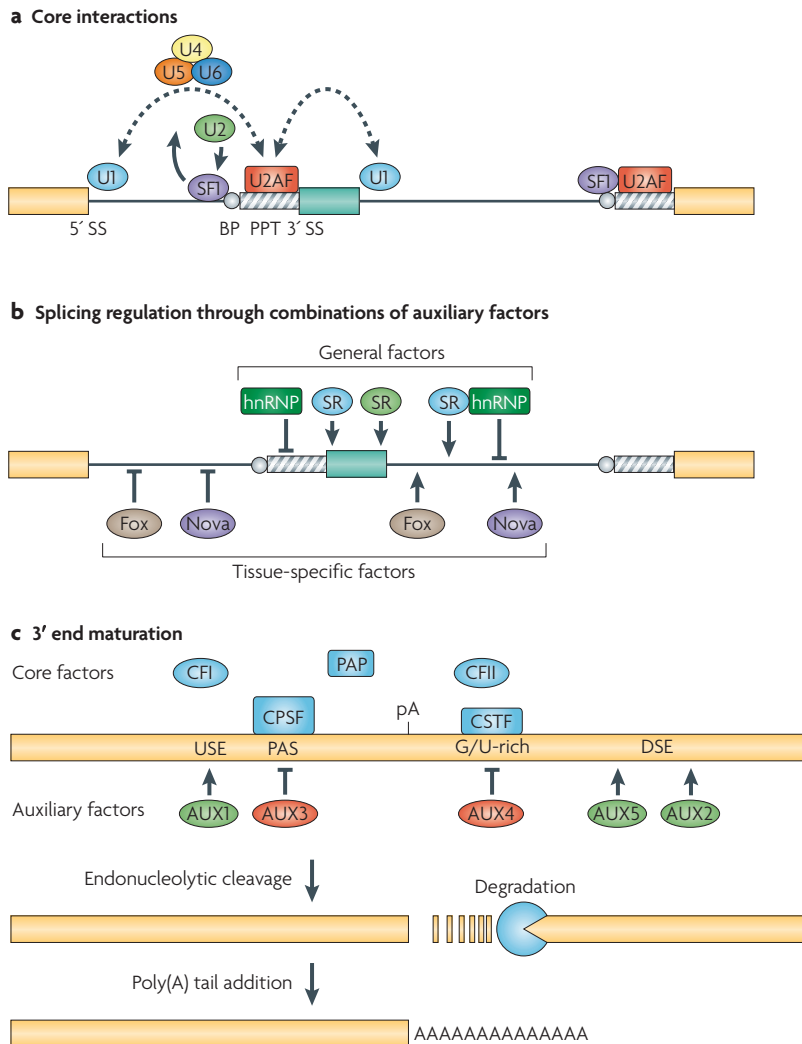


Figure 2 | Alternative splicing and polyadenylation. **a** | Core elements necessary for pre-mRNA splicing include the 5' and 3' splice sites (SS), a branch point sequence (BP) upstream of the 3' SS, and a polypyrimidine-rich tract (PPT) between the BP and the 3' SS. All of these elements are bound by components of the spliceosome, which is a dynamic macromolecular complex that consists of small nuclear RNAs (snRNAs) and ~170 proteins²⁹. 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. **b** | The combinatorial actions of both core and auxiliary splicing factors participate in the regulation of alternative splicing. For example, the serine/arginine-rich (SR) proteins comprise a family of auxiliary RNA-binding proteins (RNABPs) that bind to splicing enhancer elements to facilitate exon identification and promote splicing (although like most RNABPs they can also serve other functions in the cell). By contrast, the binding of auxiliary heterogeneous nuclear ribonucleoproteins (hnRNPs) to splicing silencer elements has a negative effect on exon inclusion; in many cases they antagonize the 'pro-splicing' activity of SR proteins. Interestingly, in addition to tissue-specific RNABPs such as Nova and Fox, the levels of some core snRNPs vary between tissues¹²⁸ and such variations might contribute to splicing regulation⁴¹. **c** | Core elements necessary for maturation of the 3' end of a mRNA include a poly(A) signal (an adenylate-rich hexameric sequence, most often AAUAAA; PAS) 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 identified⁴³, the extent to which these elements regulate alternative poly(A) site use remains unclear. AUX, auxiliary factor; CF, cleavage factor; CPSF, cleavage and polyadenylation specificity factor; CSTF, cleavage stimulation factor; DSE, downstream element; Nova, neuro-oncological ventral antigen; pA, poly(A) site; PAP, poly(A) polymerase; USE, upstream element.

New methods to analyse RNA complexity

While advances in understanding RNA regulation and complexity in the 1970s and 1980s 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 at nucleotide-level resolution. Accordingly, new methods that can simultaneously analyse multiple RNA processing events are culminating in the development of genome-wide RNA maps that pave the way for new biological insights.

Microarrays. Systematic efforts to identify RNA variants began with microarray technologies. Various different arrays have been used to elucidate RNA complexity. In particular, probe sets for alternative exons identified from genome sequencing efforts have been used to analyse splice variants. The first use of 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 spliced⁵¹. Similar exon-junction arrays have been used to identify tissue-restricted patterns of alternative mRNA expression and provide insights into their regulation by specific RNABPs⁵²⁻⁵⁴. Although microarrays provide valuable data on alternative RNA processing and diversity in different biological contexts, their use has 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 the issue of predefining targets has been the development of microarrays that can be used to interrogate 'complete' sets of transcribed exons⁵⁴. Although these arrays do not monitor specific exon junctions, they have the advantage of expanded transcriptome coverage, which provides more reliable estimates of RNA abundance. They can also detect changes in the use of individual exons (alternatively spliced isoforms) and variants derived from differential transcription regulation or alternative polyadenylation. More complete 'genome-tiling' arrays have been developed for yeast, *Drosophila melanogaster* and some human chromosomes^{55,56}; these arrays circumvent the need for prior knowledge of the transcriptome. Analyses using tiling arrays reported that most of the human genome is transcribed⁵⁷, although the biological relevance of these findings remains uncertain^{56,58}. A final limitation of microarrays is they are 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 1) takes advantage of the power of new single-molecule sequencing methods^{59,60} that can currently produce billions of nucleotides

Box 1 | RNA-seq

The term RNA-seq applies to any of several different high-throughput (next-generation) sequencing methods to obtain transcriptome-wide RNA profiles⁵⁹. Typically, RNA from two samples that are to be compared is sheared, converted to cDNA and sequenced. Using current technology this can yield up to 25 million sequence reads that are ~35 nucleotides in length⁴⁵. Although there can be sequencing bias at any particular position in the genome — for example, depending on the 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 quantified at the resolution of individual splice variants⁴⁵ or even edited RNA nucleotides⁶³. Other applications of RNA-seq using different sequencing strategies include looking at pools of RNA that are being translated by sequencing RNA bound to ribosomes¹⁵⁹ and single-cell RNA analysis¹⁶². 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 reads; the sensitivity of this technique is likely to improve over time.

Piwi-interacting RNA

A small RNA species that is processed from a single-stranded precursor RNA. They are 25–35 nucleotides in length and form complexes with the piwi protein. piRNAs are thought to have roles in transposon silencing and stem cell function.

RNA editing

The post-transcriptional modification of RNA primary sequence by the insertion and/or deletion of specific bases, or the chemical modification of adenosine to inosine or cytidine to uridine.

Exon-junction microarray

A microarray platform that contains probe sets designed to detect the mRNA sequences (junctions) formed by the splicing of one exon to another.

Ultraconserved element

A large sequence in the genome (usually >200 nucleotides) that shows high levels of conservation across multiple species.

Nonsense-mediated decay

The process by which mRNAs containing premature termination codons are destroyed to preclude the production of truncated and potentially deleterious protein products. It is also used in combination with specific alternative splicing events to control the levels of some proteins.

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 Blencowe⁶¹ and Burge⁴⁵ laboratories, who provided complete RNA profiles and analysis of alternative splicing and polyadenylation variants in different tissues that easily rivalled those that could be obtained using microarrays. The ability of RNA-seq to detect previously uncharacterized mRNA isoforms and new classes of non-coding RNAs⁶² illustrates the use 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 before sequencing. This strategy balances the constraints of hybridization technologies (as with microarrays) with the advantages of high-throughput sequencing experiments, and has been effectively used to study the RNA variants generated by RNA editing^{63,64}.

Bioinformatics. Bioinformatics has emerged as a powerful complement to current efforts to analyse 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 site⁶⁵, the poly(A) signal that is necessary for 3' end maturation and termination of transcription^{66,67}, and atypical consensus sequences that define an alternative means for regulating splicing⁶⁸. Current bioinformatic efforts are aided by and are also dependent on improvements in the number and depth of sequences available from expressed sequence tag (EST) and cDNA libraries, microarray data sets and whole-genome sequencing. Therefore, bioinformatics is likely to become more powerful as new technology improves such databases. The comparison of RNA profiles from different cell types and organisms has 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 an understanding of several 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 splicing⁷³; and defining unexpected features, such as ultraconserved elements that mediate nonsense-mediated decay (NMD) of transcripts that encode RNABPs⁷⁴. Although not the focus of this Review, bioinformatics has also been used in efforts to identify miRNA targets⁴⁷ and other regulatory elements in 3' UTRs^{38–40}.

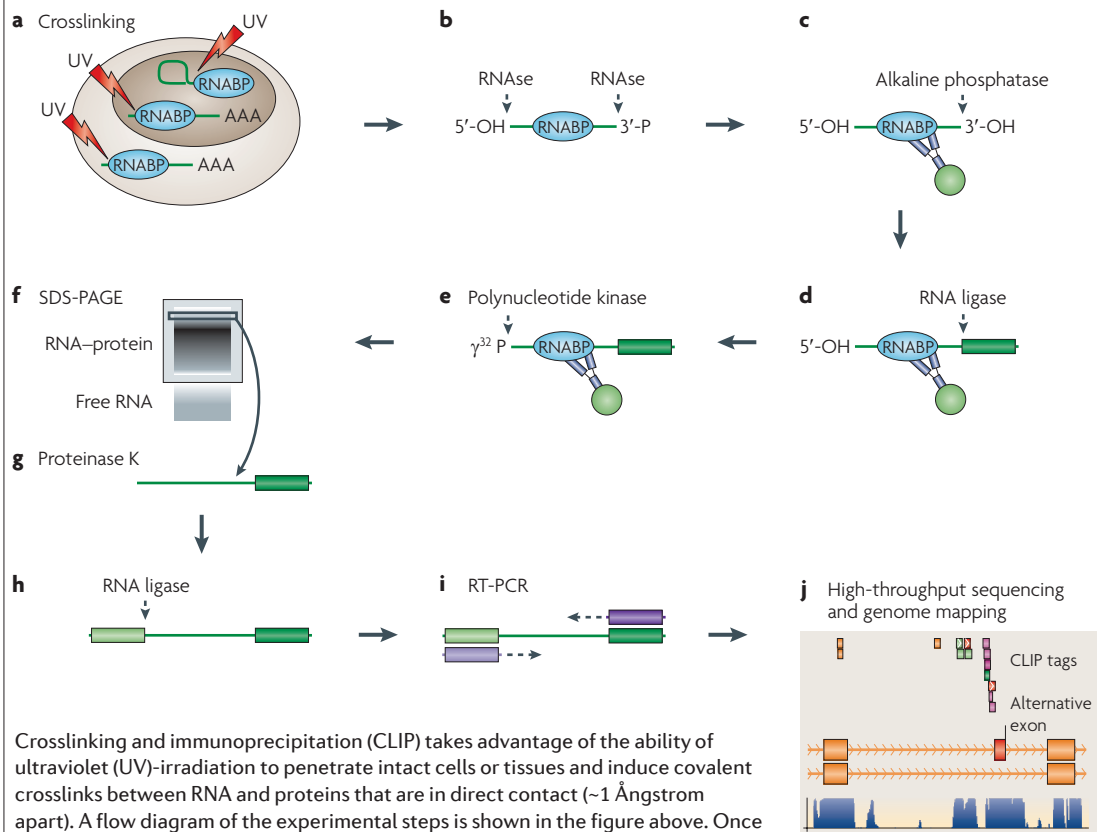
Methods to study protein–RNA interactions.

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 RNA regulation. However, these methods are limited without biochemical methods to identify the direct RNABP–RNA interactions that define RNA 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 syndrome, the loss of fragile X mental retardation 1 protein (FMRP) function is clearly the proximal 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 the RNA deregulation owing to secondary or tertiary consequences of FMRP loss⁷⁵. Put another way, any perturbation in a cell is likely to disrupt the RNA profile of that cell, as detected by methods such as microarrays or RNA-seq. Therefore, such changes in RNA profiles cannot be taken as evidence of the specific action of a RNABP. Attempts to study mechanisms of RNA regulation in cells depend on distinguishing between the direct and 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 microarrays^{76,77}. These strategies have proven useful but they cannot discriminate between direct and indirect interactions or 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 ratios, co-precipitating RNABPs and RNABP–RNA reassociation *in vitro*^{78–80}.

An alternative means of identifying regulatory RNABP–RNA interactions is the CLIP (crosslinking and immunoprecipitation) assay^{78,81,82} (BOX 2). CLIP applies the observation — first made in the study of tRNA–protein interactions in the 1970s⁸³ and even earlier for DNA–protein interactions — that ultraviolet (UV)-irradiation causes covalent crosslinking between RNA–protein complexes that are in tight apposition (that is, within ~Ångstrom distances). UV-crosslinking was applied in a cellular context in studies of protein–RNA interactions by van Venrooij⁸⁴ and Pederson^{47,85,86}

Box 2 | CLIP and HITS-CLIP methods



Crosslinking and immunoprecipitation (CLIP) takes advantage of the ability of ultraviolet (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 above. Once they have been covalently bound, RNA–protein complexes can be purified under stringent conditions, which gives the advantage of being able to separate them from closely bound RNA-binding protein (RNABP)–RNABP complexes, reassociated RNAs and background RNAs. After purification, CLIP^{81,82,92} uses proteinase K to remove the RNABP. This is followed by linker ligation and RT-PCR to analyse the RNA sequences. This sequencing analysis can be done using high-throughput sequencing methods, in which case it is referred to as ‘HITS-CLIP’ (REFS 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 analysing HITS-CLIP data can be found at [R.B.D.’s homepage](#) (see Further information box). 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 similar results^{53,88}, which suggests that crosslinking can be highly efficient across the transcriptome.

in the 1980s, and the methods were then refined by immunoprecipitation of crosslinked heterogeneous nuclear ribonucleoprotein (hnRNP)–RNA complexes by Dreyfuss⁸⁷ and colleagues.

CLIP allows the purification of RNABP–RNA interactions 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 this population provides a means of identifying the bound RNA and, importantly, the position of protein binding. CLIP (and emerging methodological improvements to this method^{81,82,88–91}), established that small, crosslinked RNA fragments could be amplified by RT-PCR after partial RNase and proteinase K digestion. This approach — initially using conventional cloning strategies⁹² 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 at 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 brain⁸⁸, stem cells⁹⁰ and tissue culture cells⁹³, and also to deconvolute Argonaute (Ago)–miRNA–mRNA ternary interactions in the mouse brain⁹¹ (discussed further below).

Biological complexity from 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 the over 36,000 sites at which RNA editing occurs⁶³. 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 (instead of DNA as the substrate, which is reviewed elsewhere^{24,26,41}) and note that, to date, most experimental validation has been achieved using protein-coding mRNAs.

Alternative splicing. Alternative splicing is one of the major ways in which RNA diversity is generated. The comparative analysis of splicing variants is yielding insights into the biological consequences of this process^{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 that expresses 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 that are necessary for neurodevelopment and the modification of synaptic strength^{95,99}.

Important general issues regarding the complexity of alternative splicing are highlighted by contrasting studies of Down syndrome cell adhesion molecule (*DSCAM*) and neurexin splicing in the nervous system. In *D. melanogaster*, *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 splicing¹⁰⁰. Despite the great complexity of RNA products and the recognition that RNABPs act to restrict *Dscam* exon usage¹⁰¹, it is believed that the choice of RNA variants produced in any one neuron is largely stochastic, and the resulting biological complexity of the RNA variants is proportionately low. Each RNA variant encodes a cell surface axonal molecule that is randomly generated to be different from that on neighbouring axons, thereby yielding a unitary outcome — that is, axon self-avoidance, which is necessary for normal fasciculation¹⁰⁰.

The regulated rather than stochastic production of alternative RNA variants has the potential to generate great diversity of biological function. The alternative splicing of neurexin pre-mRNA in mammalian brain provides an interesting example. Nearly 3,000 unique neurexin transcripts are derived from the combination of 3 genes, each of which has 2 alternate promoters and encodes transcripts with ~10 alternate exons¹⁰². This set of alternate transcripts encodes variants that give rise to alternative neurexin protein isoforms, which have different interactions with different neuroligin protein

isoforms across the synaptic cleft. This suggests that a 'splice code' might underlie trans-synaptic cell adhesion¹⁰³. There is evidence that a few RNABPs might regulate neurexin (and neuroligin) isoforms⁹⁵, suggesting that a small number of RNA regulatory proteins may generate a great diversity of biological outcomes¹⁰³. These observations also underscore the more general point that alternative splicing plays a major part in biological complexity²⁸.

Alternative polyadenylation. Although it is clear that alternative processing of pre-mRNA can confer different structural and functional properties on proteins^{76,104}, additional functional roles for alternative processing in the regulation of gene expression have also emerged. Consistent with EST-based bioinformatic studies⁴⁶, RNA-seq analysis identified tissue-specific biases in the regulation of tandem polyadenylation sites (FIGS 2, 3a). Unlike the alternative poly(A) site regulation that is coupled to the inclusion of an alternative 3' terminal exon (FIG. 1), alternative polyadenylation at tandem poly(A) sites can yield transcripts that have identical protein-coding sequences but different 3' UTR sequences. This provides the potential for differential regulation of mRNA expression by RNABPs and/or miRNAs (FIG. 3a). Exon microarray and RNA expression studies have indicated that such regulation might have important biological consequences. Proliferating cells, such as activated T lymphocytes¹⁰⁵ and tumour cells¹⁰⁶, harbour shortened 3' UTRs. By contrast, the brain, which is a non-proliferative tissue, seems to regulate polyadenylation so that transcripts harbour 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 3' UTR regulation, including miRNA-mediated regulation of translation^{105,106}, RNA localization³⁸ and RNA 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 NMD (REFS 24, 48, 107) (FIG. 3b). Although EST-based bioinformatic studies^{108,109} had suggested 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 autoregulatory manner), including serine/arginine-rich (SR) proteins⁷⁴, hnRNPs^{110–112} and core spliceosomal proteins¹⁰⁹. Interestingly, some of the exons for which splicing or skipping results in a PTC are associated with ultra-conserved 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 necessary to generate cell type- and developmental stage-specific mRNA profiles^{110,111}.

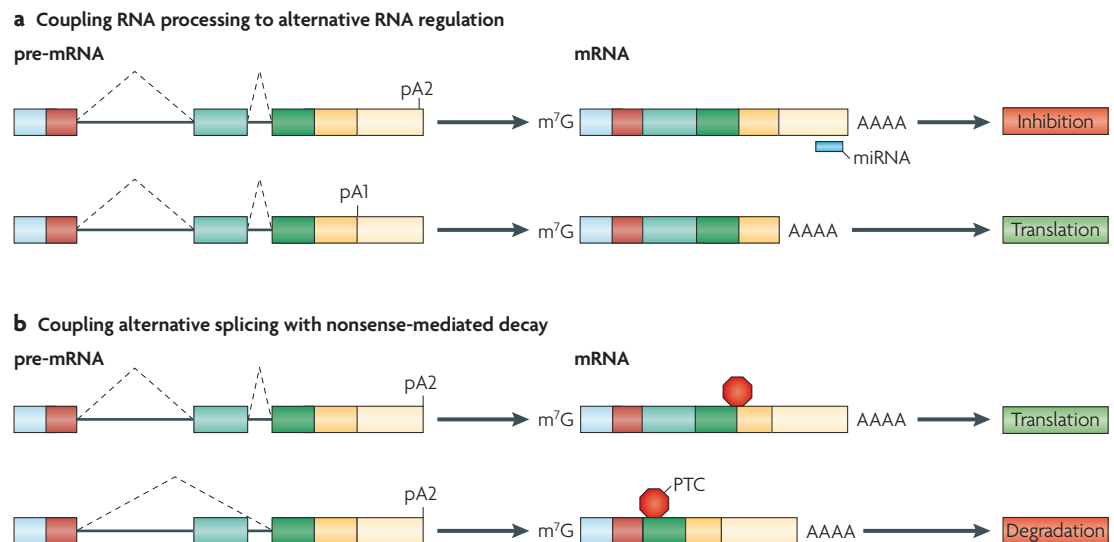


Figure 3 | Coupling of RNA processing to alternative RNA regulation. a | Alternative polyadenylation can generate mRNAs with common and isoform-specific 3' UTR sequences. Changes in 3' UTR length can alter the repertoire of regulatory elements present in the UTR, such as microRNA (miRNA) target sequences, therefore affecting the ability of the transcript to be subject to different forms of post-transcriptional regulation, in this case translation inhibition. **b** | Alternative splicing can lead to coding frameshifts, resulting in the introduction of a premature translation termination codon (PTC). The presence of a PTC triggers degradation of the mRNA by the nonsense-mediated decay pathway, therefore the regulation of alternative pre-mRNA splicing can be used to control transcript abundance, as evidenced in the physiologic regulation of the RNA-binding proteins polypyrimidine tract-binding protein 1 (PTBP1) and PTBP2 (REFS 110–112,122). m⁷G, 5' cap; pA, poly(A) site.

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 can be directly involved in the regulation of exon usage through the regulation of core factor stoichiometry^{29,114}. Similarly, SR proteins and hnRNPs are widely expressed, yet changes in their stoichiometry can mediate tissue-specific differences in alternative 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 activator¹¹⁹. Therefore, it is not sufficient to rely solely on correlative expression data to build models of RNA regulation *in vivo*.

Another 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 neuro-oncological ventral antigen (Nova) proteins *NOVA1* and *NOVA2* (REF. 78), the polypyrimidine tract-binding proteins (PTBPs)^{120–122}, embryonic lethal abnormal visual (Elav, also known as Hu) proteins¹²³, Fox proteins^{72,90}, and the CELF and muscleblind (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 crucial 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 methods 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 shown to be bound by specific RNABPs and, in most cases, *in vivo* studies have not been performed yet 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 that involves 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 mice⁹⁷, morpholino-treated zebrafish embryos¹³⁶ or cultured cells in which the expression of specific RNABPs has been knocked down by small interfering RNAs (siRNAs)^{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 is 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 in primary transcripts dictates the functional outcome of alternative pre-mRNA processing events (FIG. 4). Initial ideas relating to Nova-mediated RNA regulation in mouse brain were provided by detailed studies of two transcripts studied *in vitro* and in tissue culture cells^{134,140,141}. Subsequently, a combination of studies in Nova-knockout mice¹⁴², including exon-junction arrays⁵³, bioinformatics^{69,134} and HITS-CLIP⁸⁸, expanded these ideas into a general rule. In this work, and in subsequent studies of the FOX1/2 splicing factor that provided analogous findings^{69,72,90}, it was shown that the binding of RNABPs in an alternative exon or the flanking upstream intronic sequence is generally associated with exon skipping, whereas the binding of RNABPs to the downstream intronic sequence is generally associated with exon inclusion (FIG. 4). 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 prove to be general features of RNABP regulation, based on bioinformatic and biochemical studies of other RNABPs, including MBNL, CELF, PTBPI (REFS 69,97) and several hnRNPs (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 brain⁸⁸. Such studies illustrate a previously recognized point, albeit not made on

a transcriptome-wide scale, for SR proteins and hnRNPs: 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 *SRFS1* (previously known as ASF/SF2) in human embryonic kidney cells revealed an overrepresentation of *SFRS1* binding to mRNAs encoding RNA regulatory proteins, which suggested the possibility that a regulatory loop exists⁹³. Another new aspect of RNABP regulation emerged from CLIP analyses of *HNRNPA1*. These studies revealed that *HNRNPA1* binds to the stem–loop sequences in the miRNA precursor pre-miR-18a¹⁴³ in HeLa cells, and in so doing functions as an auxiliary factor to enhance Drosha-mediated processing to produce mature miR-18a¹⁴⁴.

Recently, HITS-CLIP was extended to the study of ternary interactions between an RNABP (Ago), RNA and miRNAs⁹¹. 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 seemed 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 by Ago HITS-CLIP maps.

RNA networks and biological coherence. Before the onset of high-throughput methods, several observations suggested that some level of biological coherence is established by RNA regulation — the idea that the coordinate regulation of RNAs that encode related proteins coordinates biological processes. Observations of biological coherence of RNA regulation during sex determination in *D. melanogaster* and iron response pathways in vertebrate cells in the 1990s 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 between direct and indirectly regulated RNAs has complicated the evaluation of such networks. Now the combination of genetic approaches, 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 (FIG. 4). For example, analysis of RNA from wild-type and Nova-knockout mouse brains using exon-junction microarrays⁵³ revealed that Nova regulates the alternative splicing of a biologically coherent set of transcripts that encode proteins with synaptic functions^{53,88,95}. HITS-CLIP and bioinformatic studies⁸⁸ showed that a subset of these transcripts were directly regulated by Nova. This network could predict aspects of Nova physiology in the mouse brain, including roles in inhibitory potentiation in the hippocampus¹⁴⁶ and in motor neuron function¹⁴⁷. Taken together, these studies provided the first demonstration in mammals of the coordinated activity of a RNABP in a biological network. Similarly, analysis of RNA regulatory defects in mouse knockouts of *Sfrs1* (REFS 135,148),

Morpholino

An oligomer of 25 nucleotides with bases linked to a morpholine ring. The oligomers can bind and inactivate selected RNA sequences on the basis of base pairing and steric interference.

Small interfering RNA

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. They 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.

Seed site

A short RNA sequence that is bound by and necessary for microRNA-mediated RNA regulation.

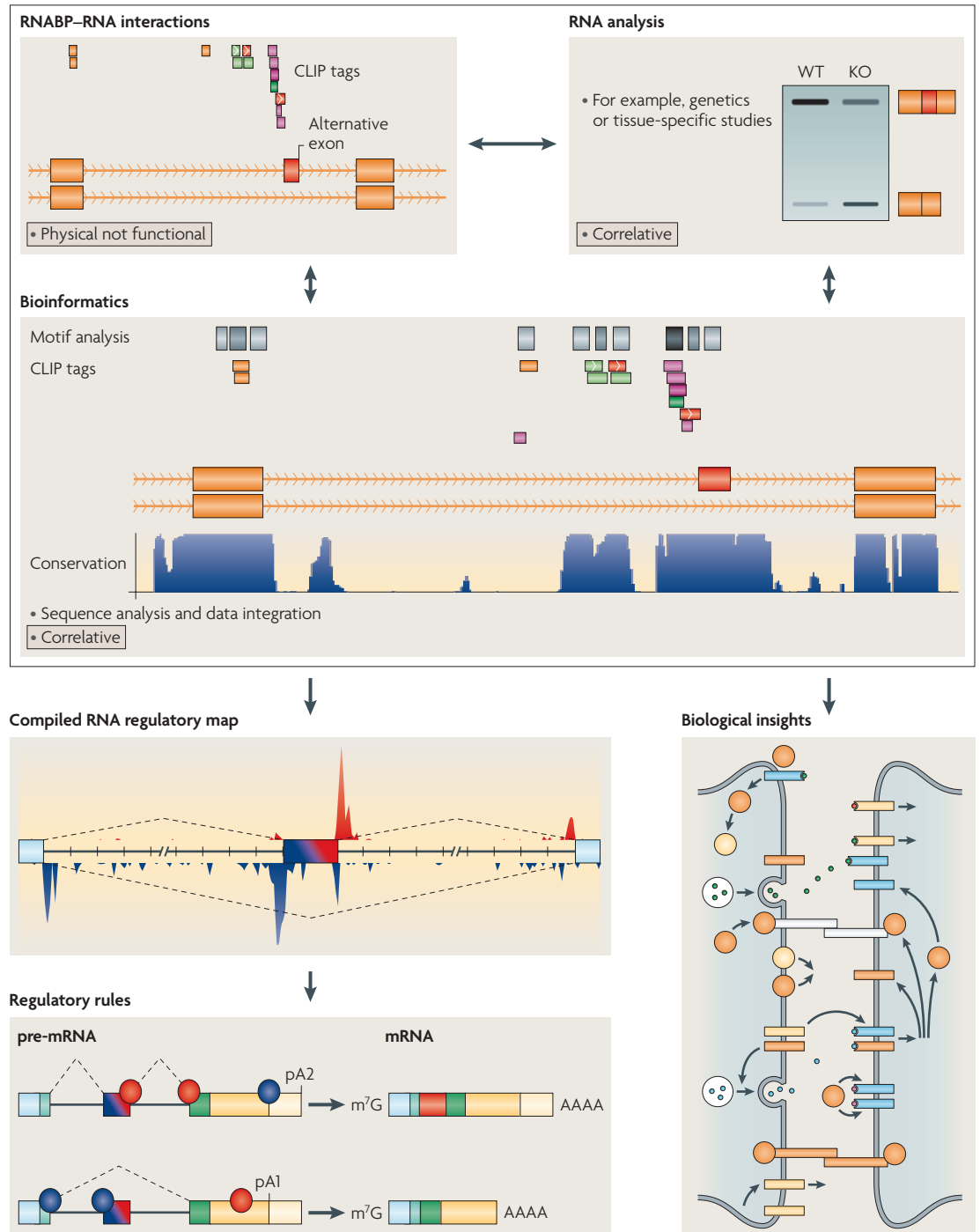


Figure 4 | Synergies between methods lead to new rules of RNA regulation. Biochemical methods as exemplified by HITS-CLIP can yield genome-wide footprints of direct RNA–protein interactions but lack functional information. By contrast, microarrays or RNA-seq can correlate differences in RNA profiles between tissues⁴⁵ or genetic systems, such as knockout (KO) and wild-type (WT) animals⁵³, but cannot distinguish direct from indirect targets. Bioinformatic analysis can also be used to identify sequence features associated with specific RNA regulatory events⁶⁹ 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 Nova¹³⁴ or FOX1/2 (REFS 69,72,90) binding enhances or inhibits the inclusion of alternative exons. m⁷G, 5' cap; pA, poly(A) site.

Srp38 (also known as *Sfrs13a*)¹³³ and *Cebpd* and *Mbnl1* (REFS 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 function.

Alternative mRNAs and disease

The importance of methods to probe mRNA complexity and understand mRNA 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. 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 seem to particularly affect complex tissues, and are exemplified by neurodegenerative disorders. RNABPs that have been linked to neurodegeneration include: FUS and TDP43, which are mutated in patients with familial amyotrophic lateral sclerosis (ALS)¹⁵⁰; Nova and the Elav proteins, which are targeted by the immune system in paraneoplastic neurodegenerative disorders^{78,151}; survival motor neuron protein 1 (SMN1), which is mutated in spinal muscular atrophy²⁷; immunoglobulin mu-binding protein 2 (IGHMBP2), which is mutated in spinal muscular atrophy and respiratory distress¹⁵²; senataxin, which is mutated in ALS4 (REF. 153); and glycyl tRNA synthetase, which is mutated in hereditary motor neuronopathy type V¹⁵⁴. Moreover, a growing number of neurological disorders is believed to be linked to RNA expansions that sequester RNABPs, as exemplified by the sequestration of MBNL by CUG repeats in myotonic dystrophy¹⁴⁹. Similarly, the deletion, mutation or inappropriate expression of miRNAs, which leads to mistargeting of the RNABP Ago and to aberrant RNA regulation¹⁷, is important in multiple disorders²⁷, including neurological 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 tumour subgroups¹⁵⁵. It seems likely that the expression of aberrantly spliced transcripts will be found to contribute to tumour 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 cancer¹⁵⁶. 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 knockdown 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 tumour suppressor¹⁵⁶ or proto-oncogene¹⁵⁷ activities might lead to 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 be likely to play a major part in advancing our understanding of disease biology and assist in the development of strategies for therapeutic intervention.

Concluding remarks and future directions

Methodological advances in the twentieth century led to the realization that RNA complexity and RNA regulation lie at the core of biological complexity. In recent years, the advent of high-throughput strategies has allowed nucleotide-level analyses of RNA regulation and complexity on a genome-wide scale. These 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 complementary methods that are described in this Review 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 (FIG. 4) 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 the 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, the activities of which are dependent on affinity constants and the 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; for example, nuclear compartments, cytoplasmic P-bodies, stress granules and dendrites, as well as the effect that the 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 to superimpose them to give a more complete picture of how RNA 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 processes¹⁵⁸.

The analysis of RNA–protein regulatory maps is also likely to yield insights into non-coding RNAs and their roles in coordinating gene regulation. Finally, the 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-seq¹⁵⁹ to characterize polyribosomal RNA, and

mouse genetics coupled to microarray profiles^{160,161} was used to profile transcribed mRNAs in 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 various cellular contexts. Such studies will enhance the discovery of how RNA regulation affects tissue complexity and disease by shaping the expression of genetic information.

- Sharp, P. A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
 - Scherrer, K., Latham, H. & Darnell, J. E. Demonstration of an unstable RNA and of a precursor to ribosomal RNA in HeLa cells. *Proc. Natl Acad. Sci. USA* **49**, 240–248 (1963).
 - Soeiro, R., Vaughan, M. H., Warner, J. R. & Darnell, J. E. J. The turnover of nuclear DNA-like RNA in HeLa cells. *J. Cell Biol.* **39**, 112–118 (1968).
 - Berk, A. J. & Sharp, P. A. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell* **12**, 721–732 (1977).
 - Berget, S. M., Moore, C. & Sharp, P. A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA* **74**, 3171–3175 (1977).
 - Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* **12**, 1–8 (1977).
 - Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
 - Breathnach, R., Mandel, J. L. & Chambon, P. Ovalbumin gene is split in chicken DNA. *Nature* **270**, 314–319 (1977).
 - Tilghman, S. M. *et al.* Intervening sequence of a mouse β -globin gene. *Proc. Natl Acad. Sci. USA* **75**, 725–729 (1978).
 - Crick, F. Split genes and RNA splicing. *Science* **204**, 264–271 (1979).
 - Darnell, J. E. J. Implications of RNA–RNA splicing in evolution of eukaryotic cells. *Science* **202**, 1257–1260 (1978).
 - Rogers, J. & Wall, R. Immunoglobulin RNA rearrangements in B lymphocyte differentiation. *Adv. Immunol.* **35**, 39–59 (1984).
 - Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* **298**, 240–244 (1982).
 - Calarco, J. A. *et al.* Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* **21**, 2963–2975 (2007).
 - Cech, T. R. Crawling out of the RNA world. *Cell* **136**, 599–602 (2009).
 - Zaug, A. J. & Cech, T. R. The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**, 470–475 (1986).
 - Carthew, R. W. & Sontheimer, E. J. Origins and mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
 - Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. *Cell* **136**, 656–668 (2009).
 - Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10**, 155–159 (2009).
 - Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
 - Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
 - Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126–130 (2008).
 - Cam, H. P., Chen, E. S. & Grewal, S. I. Transcriptional scaffolds for heterochromatin assembly. *Cell* **136**, 610–614 (2009).
 - Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688–700 (2009).
 - Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C. & Huang, T. H. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet.* **24**, 167–177 (2008).
 - Perales, R. & Bentley, D. "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* **36**, 178–191 (2009).
 - Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
 - Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243 (2002).
 - Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
 - Lutz, C. S. Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem. Biol.* **3**, 609–617 (2008).
 - Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
 - Terns, M. & Terns, R. Noncoding RNAs of the H/ACA family. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 395–405 (2006).
 - Rottman, F. M., Bokar, J. A., Narayan, P., Shambaugh, M. E. & Ludwiczak, R. N6-adenosine methylation in mRNA: substrate specificity and enzyme complexity. *Biochimie* **76**, 1109–1114 (1994).
 - Guschina, E. & Benecke, B. J. Specific and non-specific mammalian RNA terminal uridylyl transferases. *Biochim. Biophys. Acta* **1779**, 281–285 (2008).
 - Martin, G. & Keller, W. RNA-specific ribonucleotidyl transferases. *RNA* **13**, 1834–1849 (2007).
 - Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
 - Kochetov, A. V. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *Bioessays* **30**, 683–691 (2008).
 - Rodriguez, A. J., Czaplinski, K., Condeelis, J. S. & Singer, R. H. Mechanisms and cellular roles of local protein synthesis in mammalian cells. *Curr. Opin. Cell Biol.* **20**, 144–149 (2008).
 - Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
 - Richter, J. D. Breaking the code of polyadenylation-induced translation. *Cell* **132**, 335–337 (2008).
 - Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
 - 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.* **63**, 405–445 (1999).
 - Mandel, C. R., Bai, Y. & Tong, L. Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.* **65**, 1099–1122 (2008).
 - Manley, J. L. & Takagaki, Y. The end of the message – another link between yeast and mammals. *Science* **274**, 1481–1482 (1996).
 - Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
 - Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol.* **6**, R100 (2005).
 - Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
 - Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416**, 499–506 (2002).
 - Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
 - Kornblihtt, A. R. Promoter usage and alternative splicing. *Curr. Opin. Cell Biol.* **17**, 262–268 (2005).
 - Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
 - Sugnet, C. W. *et al.* Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**, e4 (2006).
 - Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nature Genet.* **37**, 844–852 (2005).
- This study shows that NOVA2 regulates the splicing of transcripts that encode a biologically coherent set of proteins with synaptic localization and function.**
- Clark, T. A. *et al.* Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* **8**, R64 (2007).
 - Johnson, J. M., Edwards, S., Shoemaker, D. & Schadt, E. E. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. *Trends Genet.* **21**, 93–102 (2005).
 - Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8**, 413–423 (2007).
 - Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
 - Seila, A. C., Core, L. J., Lis, J. T. & Sharp, P. A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564 (2009).
 - Blencowe, B. J., Ahmad, S. & Lee, L. J. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes Dev.* **23**, 1379–1386 (2009).
 - Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
 - Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
 - Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
 - Li, J. B. *et al.* Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* **324**, 1210–1213 (2009).
 - Porreca, G. J. *et al.* Multiplex amplification of large sets of human exons. *Nature Methods* **4**, 931–936 (2007).
 - Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. Are snRNPs involved in splicing? *Nature* **283**, 220–224 (1980).
 - Logan, J., Falck-Pedersen, E., Darnell, J. E. J. & 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 major globin gene. *Proc. Natl Acad. Sci. USA* **84**, 8306–8310 (1987).
 - Proudfoot, N. J. & Brownlee, G. G. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211–214 (1976).
 - Patel, A. A. & Steitz, J. A. Splicing double: insights from the second spliceosome. *Nature Rev. Mol. Cell Biol.* **4**, 960–970 (2003).
 - Castle, J. C. *et al.* Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nature Genet.* **40**, 1416–1425 (2008).
 - Sorek, R. & Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**, 1631–1637 (2003).

71. Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C. B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA* **102**, 2850–2855 (2005).
72. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
73. Ast, G. How did alternative splicing evolve? *Nature Rev. Genet.* **5**, 773–782 (2004).
74. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007). **This study reveals the presence of conserved elements and regulatory loops involved in NMD-coupled alternative splicing, which controls SR protein abundance.**
75. Darnell, J. C., Mostovetsky, O. & Darnell, R. B. FMRP RNA targets: identification and validation. *Genes Brain Behav.* **4**, 341–349 (2005).
76. Moore, M. J. & Silver, P. A. Global analysis of mRNA splicing. *RNA* **14**, 197–203 (2008).
77. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl Acad. Sci. USA* **97**, 14085–14090 (2000).
78. Darnell, R. B. Developing global insight into RNA regulation. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 321–327 (2006).
79. Kirino, Y. & Mourelatos, Z. Site-specific crosslinking of human microRNPs to RNA targets. *RNA* **14**, 2254–2259 (2008).
80. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
81. Jensen, K. B. & Darnell, R. B. CLIP: crosslinking and immunoprecipitation of *in vivo* RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **488**, 85–98 (2008).
82. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **37**, 376–386 (2005).
83. Schoemaker, H. J. & Schimmel, P. R. Photo-induced joining of a transfer RNA with its cognate aminoacyl-transfer RNA synthetase. *J. Mol. Biol.* **84**, 503–513 (1974).
84. Wagenmakers, A. J., Reinders, R. J. & van Venrooij, W. J. Cross-linking of mRNA to proteins by irradiation of intact cells with ultraviolet light. *Eur. J. Biochem.* **112**, 323–330 (1980).
85. Economidis, I. V. & Pederson, T. Structure of nuclear ribonucleoprotein: heterogeneous nuclear RNA is complexed with a major sextet of proteins *in vivo*. *Proc. Natl Acad. Sci. USA* **80**, 1599–1602 (1983).
86. Mayrand, S., Setyono, B., Greenberg, J. R. & Pederson, T. Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)⁺ heterogeneous nuclear RNA in living HeLa cells. *J. Cell Biol.* **90**, 380–384 (1981).
87. Dreyfuss, G., Choi, Y. D. & Adam, S. A. Characterization of heterogeneous nuclear RNA–protein complexes *in vivo* with monoclonal antibodies. *Mol. Cell Biol.* **4**, 1104–1114 (1984).
88. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469 (2008).
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* **48**, 287–293 (2009).
90. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nature Struct. Mol. Biol.* **16**, 130–137 (2009).
91. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature* **460**, 479–486 (2009).
92. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
93. Sanford, J. R. *et al.* Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381–394 (2009).
94. Licatalosi, D. D. & Darnell, R. B. Splicing regulation in neurologic disease. *Neuron* **52**, 93–101 (2006).
95. Ule, J. & Darnell, R. B. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.* **16**, 102–110 (2006).
96. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* **8**, 749–761 (2007).
97. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl Acad. Sci. USA* **105**, 20333–20338 (2008).
98. Xu, Q., Modrek, B. & Lee, C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* **30**, 3754–3766 (2002).
99. Li, Q., Lee, J. A. & Black, D. L. Neuronal regulation of alternative pre-mRNA splicing. *Nature Rev. Neurosci.* **8**, 819–831 (2007).
100. Hattori, D., Millard, S. S., Wojtowicz, W. M. & Zipursky, S. L. Dscam-mediated cell recognition regulates neural circuit formation. *Annu. Rev. Cell Dev. Biol.* **24**, 597–620 (2008).
101. Olson, S. *et al.* A regulator of Dscam mutually exclusive splicing fidelity. *Nature Struct. Mol. Biol.* **14**, 1134–1140 (2007).
102. Missler, M. & Sudhof, T. C. Neurexins: three genes and 1001 products. *Trends Genet.* **14**, 20–26 (1998).
103. Boucard, A. A., Chubykin, A. A., Comoletti, D., Taylor, P. & Sudhof, T. C. A splice code for *trans*-synaptic cell adhesion mediated by binding of neuroligin 1 to α - and β -neurexins. *Neuron* **48**, 229–236 (2005). **This study reveals that the interaction of proteins across the synaptic cleft is driven by specific combinations of variants generated by alternative splicing.**
104. Birzele, F., Csaba, G. & Zimmer, R. Alternative splicing and protein structure evolution. *Nucleic Acids Res.* **36**, 550–558 (2008).
105. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008). **This study shows that mRNA 3' UTR length can be dynamically regulated in response to cell stimulation, allowing the escape of mRNAs from post-transcriptional control.**
106. Mayr, C. & Bartel, D. P. Widespread shortening of 3' UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
107. Lejeune, F. & Maquat, L. E. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.* **17**, 309–315 (2005).
108. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100**, 189–192 (2003).
109. Saltzman, A. L. *et al.* Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell Biol.* **28**, 4320–4330 (2008).
110. Boutz, P. L. *et al.* A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* **21**, 1636–1652 (2007).
111. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**, 435–448 (2007).
112. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**, 91–100 (2004).
113. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
114. Yu, Y. *et al.* Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. *Cell* **135**, 1224–1236 (2008).
115. Caceres, J. F., Stamm, S., Helfan, D. M. & Krainer, A. R. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* **265**, 1706–1709 (1994).
116. Ge, H. & Manley, J. L. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell* **62**, 25–34 (1990).
117. Krainer, A. R., Conway, G. C. & Kozak, D. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35–42 (1990).
118. Biamonti, G. & Caceres, J. F. Cellular stress and RNA splicing. *Trends Biochem. Sci.* **34**, 146–153 (2009).
119. Feng, Y., Chen, M. & Manley, J. L. Phosphorylation switches the general splicing repressor SRp58 to a sequence-specific activator. *Nature Struct. Mol. Biol.* **15**, 1040–1048 (2008). **This study shows how post-translational regulation of a RNABP can alter its functional properties.**
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.* **20**, 7463–7479 (2000).
121. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97**, 6350–6355 (2000).
122. Spellman, R., Llorian, M. & Smith, C. W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **27**, 420–434 (2007).
123. Okano, H. J. & Darnell, R. B. A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J. Neurosci.* **17**, 3024–3037 (1997).
124. Vicente, M. *et al.* Muscleblind isoforms are functionally distinct and regulate α -actinin splicing. *Differentiation* **75**, 427–440 (2007).
125. Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
126. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008). **This study uses RNA-seq and bioinformatics analyses to investigate potential mechanisms by which RNA processing events are coordinately regulated in different tissues.**
127. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
128. Grosso, A. R. *et al.* Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res.* **36**, 4823–4832 (2008).
129. Das, D. *et al.* A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. *Nucleic Acids Res.* **35**, 4845–4857 (2007).
130. Fagnani, M. *et al.* Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol.* **8**, R108 (2007).
131. Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**, 1485–1493 (2005).
132. Jelen, N., Ule, J., Zivin, M. & Darnell, R. B. Evolution of Nova-dependent splicing regulation in the brain. *PLoS Genet.* **3**, 1838–1847 (2007).
133. Feng, Y. *et al.* SRp38 regulates alternative splicing and is required for Ca²⁺ handling in the embryonic heart. *Dev. Cell* **16**, 528–538 (2009).
134. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580–586 (2006).
135. Xu, X. *et al.* ASF/SF2-regulated CaMKII β alternative splicing temporally reprograms excitation–contraction coupling in cardiac muscle. *Cell* **120**, 59–72 (2005).
136. Calarco, J. A. *et al.* Regulation of vertebrate nervous system-specific alternative splicing and development by an SR-related protein. *Cell* **138**, 898–910 (2009).
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.* **18**, 1247–1258 (2008).
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* **33**, 438–449 (2009).
139. Venables, J. P. *et al.* Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell Biol.* **28**, 6033–6043 (2008).
140. Dredge, B. K. & Darnell, R. B. Nova regulates GABA_A receptor $\gamma 2$ alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol. Cell Biol.* **23**, 4687–4700 (2003).
141. Dredge, B. K., Stefani, G., Engelhard, C. C. & Darnell, R. B. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24**, 1608–1620 (2005).

142. Jensen, K. B. *et al.* Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25**, 359–371 (2000).
143. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
144. Michlewski, G., Guil, S., Semple, C. A. & Caceres, J. F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–393 (2008).
145. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).
146. Huang, C. S. *et al.* Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. *Cell* **123**, 105–118 (2005).
147. Ruggiu, M. *et al.* Rescuing Z⁺ agrin splicing in *Nova* null mice restores synapse formation and unmask a physiologic defect in motor neuron firing. *Proc. Natl Acad. Sci. USA* **106**, 3513–3518 (2009).
148. Kanadia, R. N., Clark, V. E., Punzo, C., Trimarchi, J. M. & Cepko, C. L. Temporal requirement of the alternative splicing factor Sfrs1 for the survival of retinal neurons. *Development* **135**, 3923–3933 (2008).
149. O'Rourke, J. R. & Swanson, M. S. Mechanisms of RNA-mediated disease. *J. Biol. Chem.* **284**, 7419–7423 (2009).
150. Lagier-Tourenne, C. & Cleveland, D. W. Rethinking ALS: the FUS about TDP-43. *Cell* **136**, 1001–1004 (2009).
151. Darnell, R. B. & Posner, J. B. Paraneoplastic syndromes involving the nervous system. *N. Engl. J. Med.* **349**, 1543–1554 (2003).
152. Grohmann, K. *et al.* Mutations in the gene encoding immunoglobulin μ -binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nature Genet.* **29**, 75–77 (2001).
153. Chen, Y. Z. *et al.* DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74**, 1128–1135 (2004).
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.* **72**, 1293–1299 (2003).
155. Grosso, A. R., Martins, S. & Carmo-Fonseca, M. The emerging role of splicing factors in cancer. *EMBO Rep.* **9**, 1087–1093 (2008).
156. Venables, J. P. *et al.* Cancer-associated regulation of alternative splicing. *Nature Struct. Mol. Biol.* **16**, 670–676 (2009).
- This study combined bioinformatics and functional studies to establish a link between splicing deregulation and cancer.**
157. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature Struct. Mol. Biol.* **14**, 185–193 (2007).
158. Richter, J. D. & Klann, E. Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes Dev.* **23**, 1–11 (2009).
159. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
160. Doyle, J. P. *et al.* Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
161. Heiman, M. *et al.* A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
162. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382 (2009).

Acknowledgements

We are grateful to members of the laboratory for thoughtful discussions. We apologize to the many colleagues whose interesting studies we reviewed but were unable to cite here due to space limitations. This work was supported by grants from the National Institute of Health and the Howard Hughes Medical Institute.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

UniProtKB: <http://www.uniprot.org>
 DSCAM | EMRP | EOX2 | HNRNPA1 | NOVA1 | NOVA2 | PTBP1 | PTBP2 | SRSF1

FURTHER INFORMATION

Robert Darnell's homepage: <http://www.rockefeller.edu/labheads/darnellr>
 Nature Reviews Genetics article series on Applications of Next-generation Sequencing: <http://www.nature.com/nrg/series/nextgeneration/index.html>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF