Review

CellPress

RNA-binding proteins in neurodegeneration: Seq and you shall receive

Julia K. Nussbacher[1], Ranjan Batra[1], Clotilde Lagier-Tourenne[2,3], and Gene W. Yeo[1,4]

[1] Department of Cellular and Molecule Medicine, Institute for Genomic Medicine, UCSD Stem Cell Program, University of California, San Diego, La Jolla, CA, USA
[2] Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA
[3] Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, CA, USA
[4] Department of Physiology, National University of Singapore, Singapore

As critical players in gene regulation, RNA binding proteins (RBPs) are taking center stage in our understanding of cellular function and disease. In our era of bench-top sequencers and unprecedented computational power, biological questions can be addressed in a systematic, genome-wide manner. Development of high-throughput sequencing (Seq) methodologies provides unparalleled potential to discover new mechanisms of disease-associated perturbations of RNA homeostasis. Complementary to candidate single-gene studies, these innovative technologies may elicit the discovery of unexpected mechanisms, and enable us to determine the widespread influence of the multifunctional RBPs on their targets. Given that the disruption of RNA processing is increasingly implicated in neurological diseases, these approaches will continue to provide insights into the roles of RBPs in disease pathogenesis.

RBPs and RNA processing

If DNA is the blueprint for a cell, then transcribed RNA represents bits of information retrieved from DNA to direct cellular function and promote cell survival. Before guiding cell function, these nascent RNAs must first undergo extensive processing and precise localization, both of which are dynamic processes that require complex interplay among proteins interacting with RNA, known as RBPs (see Glossary). As with any multistep, multicomponent procedure, exact homeostatic control of RNA processing is essential for the sustained health and proper function of the eukaryotic cell. RBPs bind to specific sequences or secondary structures within the RNA molecule to modulate co- and post-transcriptional processing steps (Figure 1).

Opportunities for misregulation of RNA processing abound, often caused by mutations within RBP binding sites or in the RBPs themselves, altering RBP–RNA interactions. Such dysfunction has been identified as the culprit

Corresponding authors: Lagier-Tourenne, C. (clagiert@ucsd.edu); Yeo, G.W. (geneyeo@ucsd.edu).
Keywords: RNA binding protein; neurodegeneration; high-throughput sequencing; splicing; polyadenylation.

0166-2236/
 http://dx.doi.org/10.1016/j.tins.2015.02.003

Glossary

Crosslink: formation of a covalent bond between two entities. In the context of this review, crosslinking refers to the formation of covalent bonds between protein and nucleic acid that are within close physical proximity (within a few angstroms). They can be chemical and reversible (formaldehyde), or photochemical and irreversible (UV light).
Library: the pooled sample of fragmented nucleic acids having the necessary adapters for high-throughput sequencing.
Polyadenylation: the process of adding multiple adenosine residues to the 3′ end of transcripts. The poly(A) tail is necessary for nuclear export as well as protecting the 3′ end of the transcript from exonuclease degradation. Poly(A) sites can be located within introns, exons or the 3′UTR of a transcript; however, poly(A) sites in the 3′UTR are more commonly utilized *in vivo*. Alternative polyadenylation is a common phenomenon in which one of many potential poly(A) sites available is favored. Use of a poly(A) site depends on the core 3′-processing machinery, the strength of *cis*-elements, transcription dynamics, and other auxiliary factors [119].
Randomer: for a defined length of nucleic acid, the set of oligomers with all possible sequences.
Read-mapping: the process of aligning short sequencing reads to a reference genome or sequence.
RNA-binding protein (RBP): a protein that interacts with RNA to affect downstream function or processing.
RNA element: sequence of RNA that is often conserved and has a particular function, for example as a binding site for an RBP.
RNA splicing: the process of excising non protein-coding regions of pre-mRNA, called introns, and the joining of exons. The preferential inclusion or exclusion of an exon is termed 'alternative splicing' and contributes significantly to the diversity of the proteome.
RNA splint ligation: the ligation of two RNA molecules brought together via binding of a third bridging oligonucleotide complementary to the two RNA molecules.
RNA turnover: the process of RNA degradation. There are several known mechanisms, all of which involve the recruitment and function of several RBPs [120]. Most RNAs are degraded in a deadenylation-independent manner, in which the poly(A) tail is shortened, followed by removal of the 5′ cap, enabling exonuclease degradation of the RNA. Transcripts can also be targeted for degradation without deadenylation or decapping via miRNA-mediated recruitment of the RNA-induced silencing complex (RISC) complex. Another deadenylation-independent mechanism of RNA turnover is nonsense-mediated decay (NMD), where the interaction of RBPs Upf1, Upf3, and Nmd2 with mRNAs that contain premature stop codons results in decapping and degradation by exonucleases [121]. RNAs lacking a stop codon are also rapidly deadenylated and subjected to decapping and exonuclease degradation in a pathway known as 'nonstop decay' [122,123].
Sequencing adapter: defined nucleic acid sequences ligated to the end of the nucleic acid fragments of interest before sequencing; enables hybridization to a sequencing flow cell as well as recognition by the sequencing primer.
Untranslated region (UTR): the regions at the 5′ and 3′ ends of transcripts that do not encode protein, but often harbor *cis*-regulatory elements that are bound by protein.

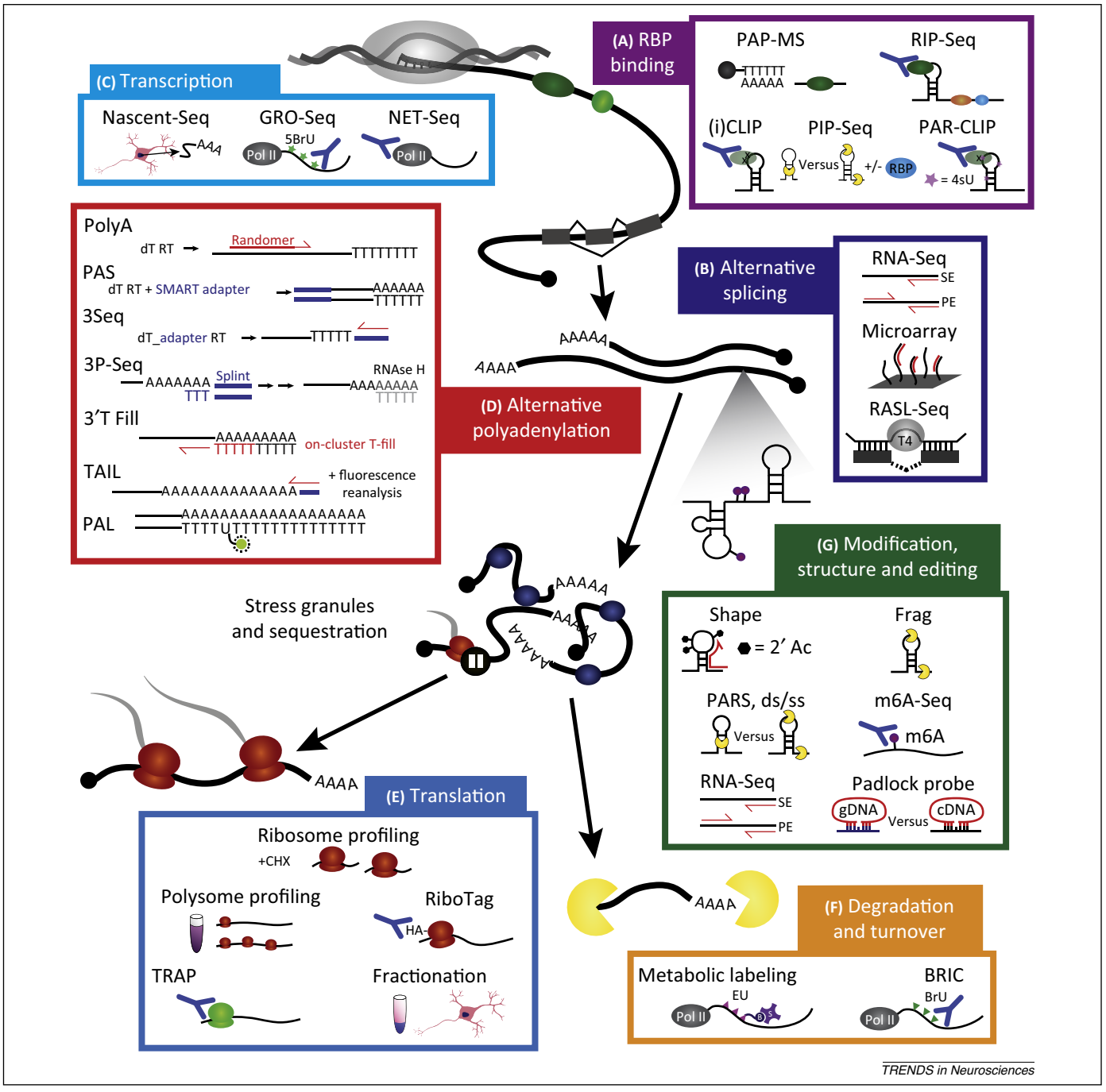

Figure 1. High-throughput sequencing (Seq) enables quantification of RNA processing steps on a global scale. **(A)** RNA-binding protein (RBP) binding. Poly(A) purification and liquid chromatography mass spectrometry (PAP-MS) is a method that involves poly(A) selection of RBP-bound RNA followed by proteomic analysis to identify mRNA-bound proteins, enabling the identification of novel RBPs. RNA immunoprecipitation (RIP-Seq) is a method for identifying whole transcripts associated with an RBP by immunoprecipitating the RBP with bound RNA, then subjecting the isolated RNA to RNA-Seq analysis. Crosslinking immunoprecipitation (CLIP)-Seq and individual nucleotide resolution CLIP (iCLIP) improve on the resolution of RIP-Seq by utilizing UV light to crosslink RNAs to protein. This enables both more stringent washing to reduce false positives, and a digestion step that reveals specific RBP target regions and motifs. Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) is similar to iCLIP, but crosslinking efficiency is increased with the metabolic labeling of RNA by 4-thio-uracil (4sU). Protein interaction profile (PIP)-Seq characterizes the structural dependence of RBP–RNA interactions through the inclusion of proteinase K-dependent libraries. **(B)** Alternative splicing. Splicing inclusion and exclusion events can be detected by a simple RNA-Seq experiment, either single-end (SE) or paired-end (PE) reads, provided there is sufficient read depth at exon–exon and exon–intron junctions, and does not require the splicing event to be annotated. Microarrays can also be used to detect splicing changes, and are more sensitive to detecting events in lowly expressed transcripts compared with RNA-Seq; however, the assay is limited to the number of events on an array, as well as prior knowledge of a splicing event. A less comprehensive but sensitive technique, RNA-mediated oligonucleotide Annealing, Selection, and Ligation with sequencing (RASL-Seq), utilizes a ligation reaction to detect an event based on ligation of oligomers complementary to alternative exon–exon junctions. **(C)** Transcription. Nascent-Seq involves the sequencing of nascent RNAs isolated from the nucleus using centrifugation and fractionation. Global run-on sequencing (GRO-Seq) involves pausing of the transcription machinery, then reinitiation with the addition of brominated nucleotides, which are incorporated into nascent transcripts and facilitate immunopurification of the nascent RNAs by antibodies specific for 5-bromouridine (5BrU). To avoid transcription pausing, native elongating transcript sequencing (NET-Seq) uses immunopurification of RNA polymerase II (pol II) with its associated transcripts for sequencing. **(D)** Alternative polyadenylation. Poly(A) site usage can be determined by several techniques, the majority of which involve positive selection by oligo(d)Ts and sequencing into the poly(A) tail. Poly(A)-Seq involves oligo(d)T-primed first strand synthesis and randomer-primed second strand synthesis. Poly(A) site sequencing (PAS-Seq) also uses oligo(d)T-primed first strand synthesis, but with the inclusion of a SMART adapter added at the end of first strand synthesis, reducing the need for internal randomer priming. 3Seq takes a somewhat modified approach, with oligo(d)T-primed first strand synthesis that includes an adapter for second strand synthesis. A problem with internal priming is the risk that the PCR will not proceed all the way to the

(Figure legend continued on the bottom of the next page.)

in countless human diseases [1,2] and is increasingly recognized as a central component in neurodegenerative disorders. In mice, more than half of the known or putative RBPs can be detected in brain tissue by *in situ* hybridization, and a subset of these are specific to neural cells [3], consistent with neurons being susceptible to mutations in RBP binding elements and aberrant RBP interactions. Furthermore, RBPs expressed in the central nervous system are intimately involved in the regulation of alternative splicing, which is also more prevalent in cells of the nervous system than in any other cell type [4,5]. Finally, RBPs are required to protect mRNAs from premature translation and degradation during their transport from soma to dendrites and axons, enabling *de novo* protein synthesis at synapses [6–9]. Genomic approaches have recently provided major insights into the multiple ways in which RBPs influence the fate of their targets and create vast opportunities to reveal the roles of RBPs in neurological diseases. In this review, we describe genome-wide technologies used to identify and characterize RBPs in the context of RNA processing and neurological disease.

Identifying RBPs and their targets

Although hundreds of RBPs have been predicted based on homology to known RNA-binding domains, only a subset of these have been validated and characterized *in vivo*. To identify novel RBPs, poly(A) affinity purification and mass spectrometry (PAP-MS) is a straightforward technique in which mRNA–protein complexes are purified by poly(A) selection, and bound proteins are identified by mass spectrometry (Figure 1A). This technique has enabled identification of RBPs lacking a canonical RNA-binding domain [10]; however, poly(A) selection misses intron-bound RBPs and RBPs bound to nonpolyadenylated species, such as unprocessed mRNA, miRNAs, and their precursors. Putative RBPs can then be validated by techniques utilizing immunoprecipitation and Seq. The most basic of these is RNA immunoprecipitation (RIP)-Seq [11] (Figure 1A). This technique identifies RBP-associated transcripts, but does not reveal precise binding sites, and is potentially encumbered by false positives due to low stringency washes in the absence of cross-linking [12]. To determine specific binding sites, cross-linking and immunoprecipitation (CLIP)-Seq (or HITS-CLIP) [15,25,28] is now commonly used (Figure 1A). UV irradiation (inefficiently) induces covalent bonds (crosslinks) between proteins and nucleic acids, enabling both stringent washes to remove nonspecific and indirect interactions, as well as RNA size-trimming by RNase digestion to hone in on specific binding sites [14]. CLIP-Seq has proved invaluable for precise identification of *in vivo* RBP-binding sites and providing insights into RBP functions in disease and development [15–29]). Variations of CLIP-Seq include metabolic labeling with photoreactive thiolated nucleotides to enhance UV-crosslinking efficiency (Figure 1A). This technique, termed 'photoactivatable-ribonucleoside-enhanced CLIP' (PAR-CLIP) [30] gets closer to nucleotide-level resolution of binding sites [31], with the caveat that 4-thiouridine at high concentrations was recently shown to inhibit rRNA synthesis and induce a nucleolar stress response [32]. Leveraging the observation that reverse transcription of isolated RNA terminates at the crosslinked nucleotide, iCLIP [33] and crosslink-induced mutation sites (CIMS) [13,124] (Figure 1A) pinpoint the exact site of protein–RNA interaction. CLIP data can also be analyzed in conjunction with *in vitro* techniques, such as RNA SELEX [34], SEQRS [35], RNAcompete [36,37], RNA Bind-n-Seq [38], or interactome profiling [39,125].

A disadvantage of the CLIP methods is their inability to identify a binding site comprising multiple motifs that are distal in the RNA primary sequence, but form a single binding site through secondary structure formation, such as a stem-loop [60]. Although RBPs often have primary sequence specificities, it is likely that they recognize these sequences in the context of a particular RNA structure. CLIP methods are limited by their reliance on RNA–protein crosslink formation, whose efficiency is dependent on the molecular geometry of the RNA–protein interface [40]. Alternative techniques have been developed, such as protein interaction profile sequencing (PIP-Seq) [41] (Figure 1A), which utilizes single- and double-strand-specific RNases, together with or without proteinase to uncover how RNA structure influences RBP binding. Computational approaches may also aid in the prediction of the RNA structure at RBP binding sites, as was done with the amyotrophic lateral sclerosis (ALS)-associated FUS/TLS [42] and Lin28 [43,44].

poly(A) site, making it impossible to determine alternative polyadenylation. Furthermore, oligo(d)T priming risks polymerase slippage. 3P-Seq avoids this with the use of a splint ligation to attach adapters, while 3′T-fill fills in the poly(A) tract with d(T)s on the flow cell of the sequencing instrument before initiation of sequencing. In addition to alternative poly(A) site identification, poly(A) length can be critical in the case of RBPs that bind the poly(A) tail. Poly(A) size can be determined by TAIL-Seq, which utilizes special software to reanalyze the fluorescence signal on the flow cell to eliminate false 'T' base calls due to residual signal from reading the poly(A) tract. Alternatively, poly(A) tail length sequencing (PAL-Seq) utilizes stoichiometric incorporation of a modified uridine during poly(A) sequencing that can be fluorescently tagged, identifying tail length as proportional to fluorescence intensity. **(E)** Translation. Methods for identifying actively translating transcripts involve purification of ribosomes or polysomes by fractionation and sequencing of the associated RNAs, as in ribosome and polysome profiling. Ribosome labeling has enhanced these techniques. In a method similar to CLIP, tagged ribosomes can be immunoprecipitated and the associated RNAs sequenced, as in RiboTag. Alternatively, lineage-specific promoters driving EGFP-labeled ribosomes enable the cell-specific isolation of polysomes, as in TRAP-Seq, further enhancing the resolution of nascent RNA sequencing. One aspect of translational control is localization of the translation machinery and associated RNAs, which can be probed by compartmental fractionation and either sequencing or proteomic analysis of each fraction. **(F)** Degradation and turnover. Measuring rates of global mRNA synthesis and degradation is most commonly carried out with a pulse-chase experiment to label nascent transcripts followed by the observed loss of label as transcripts are degraded. Metabolic labeling can be performed with 5-ethynyl uridine (EU), which can be biotinylated with click chemistry for purification of pulse-labeled transcripts, or with 4-thiouridine, which can also be biotinylated. Similarly, RNA can be pulsed with 5BrU in BrU immunoprecipitation chase-deep sequencing analysis (BRIC-Seq). These techniques require an immense amount of input material, making it a challenging assay for difficult-to-obtain cell types, such as iPSC-derived neurons. **(G)** Modifications, structure & editing. High-throughput methods for determining secondary structure include selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)-Seq, which involves the acetylation of specific bases in single-stranded loops and bulges, blocking cDNA synthesis and enabling single-nucleotide resolution of certain secondary structures. To glean information on regions of secondary structure, differential digestion with single- and double-strand-specific nucleases can be utilized. Fragmentation sequencing (Frag-Seq) utilizes a single-strand-specific nuclease to identify regions of RNA involved in base-pairing structures. Parallel analysis of RNA structure (PARS)-Seq compares sequencing libraries of single-strand and double-strand-specific nuclease-treated RNA to identify secondary structure. To identify RNA editing events, such as A-to-I editing, RNA-Seq often suffices. However, to probe a specific site, padlock probes can be designed to compare the identity of a single base between cDNA generated from RNA to the genomic DNA. m6A-Seq enables low-resolution sequencing of m6A modifications through nucleobase IP.

CLIP-Seq has also provided valuable insights into the roles of RBPs in neural development and disease. Genome-wide identification of RBP-binding sites was achieved for the neuron-specific Nova proteins involved in paraneoplastic opsoclonus-myoclonus ataxia (POMA) [15]. Thousands of binding sites and functional rules underlying Nova regulation of splicing were identified by correlating binding sites with splicing alterations induced by the absence of Nova. Similar efforts in cultured cells [18,19,42,45–47] or mouse and human brains [26,27,48,49] have demonstrated largely different binding patterns for TDP-43 and FUS/TLS, two RBPs linked to ALS/frontotemporal dementia (FTD). CLIP-Seq of the muscleblind proteins involved in myotonic dystrophy (DM) [17,22,50] identified mostly 3′ untranslated region (UTR) and intronic binding, supporting a role in regulation of splicing as well as subcellular localization and translation for a subset of targets. CLIP-Seq of the cytoplasmic polyribosome-associated fragile X mental retardation protein (FMRP) involved in fragile X syndrome (FXS) also revealed a distinct binding pattern through coding regions of its RNA targets, consistent with a role for FMRP in translational repression by promoting ribosome stalling on mRNAs [20]. Finally, CLIP has also suggested novel RNA targets for the protein Park7 (DJ-1) involved in early-onset Parkinson's disease [51,52]. Overall, CLIP-Seq combined with appropriate computational analyses has become a powerful tool for elucidating RBP functions and for providing significant insight into the mechanisms by which misregulation of these RBPs leads to neurodegenerative disease.

Exploring alternative splicing co- and post-transcriptionally

Pre-mRNA splicing, the process of intron removal and joining of exons, is tightly regulated by RBPs, several of which, including MBNL1/2, TDP-43, FUS/TLS, TAF15, EWS, hnRNPA1 and hnRNPA2/B1, have been implicated in neurodegenerative diseases, such as myotonic dystrophy, multisystem proteinopathy, and ALS. The potential impact of alternative splicing, the process whereby multiple isoforms are generated from the same genic locus [53,54], is also increasingly recognized in neurodegenerative diseases, including Alzheimer's disease and Parkinson's disease [55] (Figure 2). Over the past decade, microarray and sequencing studies have revealed that >90% of human genes undergo alternative splicing. Expectedly, disrupting the function of a single RBP often results in a dramatic effect on transcript diversity [15,25–27]. To identify alternatively spliced exons, multiple high-throughput techniques have been developed (Figure 1B). An early technique utilized microarrays that interrogate exon–exon junctions [56,57]. However, this method is limited by the number of probes that can fit on an array, and only examines already annotated events. As an alternative approach, both novel and annotated splicing events can be identified by standard RNA-Seq through the analysis of exon–junction reads [58–60,126], but sensitivity is dependent on sequencing depth (i.e., the number of reads that map to a particular genic region). Given that sequencing depth is proportional to cost, this approach is not cost-effective but allows for novel isoform discovery. Nevertheless, in the case of lowly expressed alternative splicing events, microarrays can outperform RNA-Seq. Another sequencing-based, targeted approach to measuring alternatively spliced events is RNA-mediated oligonucleotide Annealing, Selection, and Ligation with next-generation sequencing (RASL-Seq) [61,62]. Here, for each exon–exon junction, a pair of DNA oligonucleotides is designed to hybridize immediately upstream and downstream of the junction. Following annealing, mRNA is captured on a poly(A)-selective solid support and unbound oligonucleotide probes are removed. Treatment with ligase joins the pair of DNA probes to form a PCR-amplifiable product only in the presence of the exon–exon junction. Using pools of DNA probe pairs during ligation and barcoded (sequence-indexed) primers during PCR enables the simultaneous interrogation of several hundred to thousands of specific splicing events from a large number of samples in a single next-generation sequencing run, thus making this method amenable for large-scale drug screening [62].

Splicing of many human genes has previously been shown to be co-transcriptional [63,64]. Intriguingly, a class of RBPs known as the FET family (comprising FUS/TLS, EWS and TAF15) has been associated with ALS/FTD and proposed to affect both transcription elongation and splicing [18,26,48,65–67] (Figure 2). To assess co-transcriptional splicing, several methods have been developed. Nascent-Seq (Figure 1C) utilizes subcellular fractionation to isolate nascent transcripts for sequencing [68], enabling the genome-wide study of co-transcriptional regulation mechanisms. Several variations of this technique have emerged (Figure 1C) including Global Run-On sequencing (GRO-Seq), where transcription initiation and elongation are halted and restarted in the presence of the nucleotide analog 5-bromouridine 5′-triphosphate (BrUTP). Nascent RNA is then isolated by BrUTP immunoprecipitation and sequenced [69], enabling identification of actively transcribed regions, but requiring reinitiation of transcription elongation under artificial conditions. By contrast, Native Elongating Transcript sequencing (NET-Seq) identifies nascent transcripts under physiological conditions [70] via immunoprecipitation of the polymerase (pol) II complex and associated RNAs without crosslinking (Figure 1C). GRO-Seq and NET-Seq can be directly applied to study *in vivo* nascent RNA populations. DM, ALS/FTD, spinal muscular atrophy (SMA), and POMA are all conditions accompanied by splicing alterations due to the disruption of different RBPs, yet it is not clear whether these RBP-regulated splicing events occur co- or post-transcriptionally. Enlisting these techniques to quantify splicing changes in healthy and patient tissues as well as cellular or animal models of disease may uncover interplays between elongation and disease-linked splicing alterations and provide insight into subtle differences between the effects of various mutants that might correlate with therapeutic sensitivity.

Identifying alternative polyadenylation by PolyA sequencing

The selection of the 3′ end cleavage site within pre-mRNA transcripts followed by addition of a poly(A) tail is

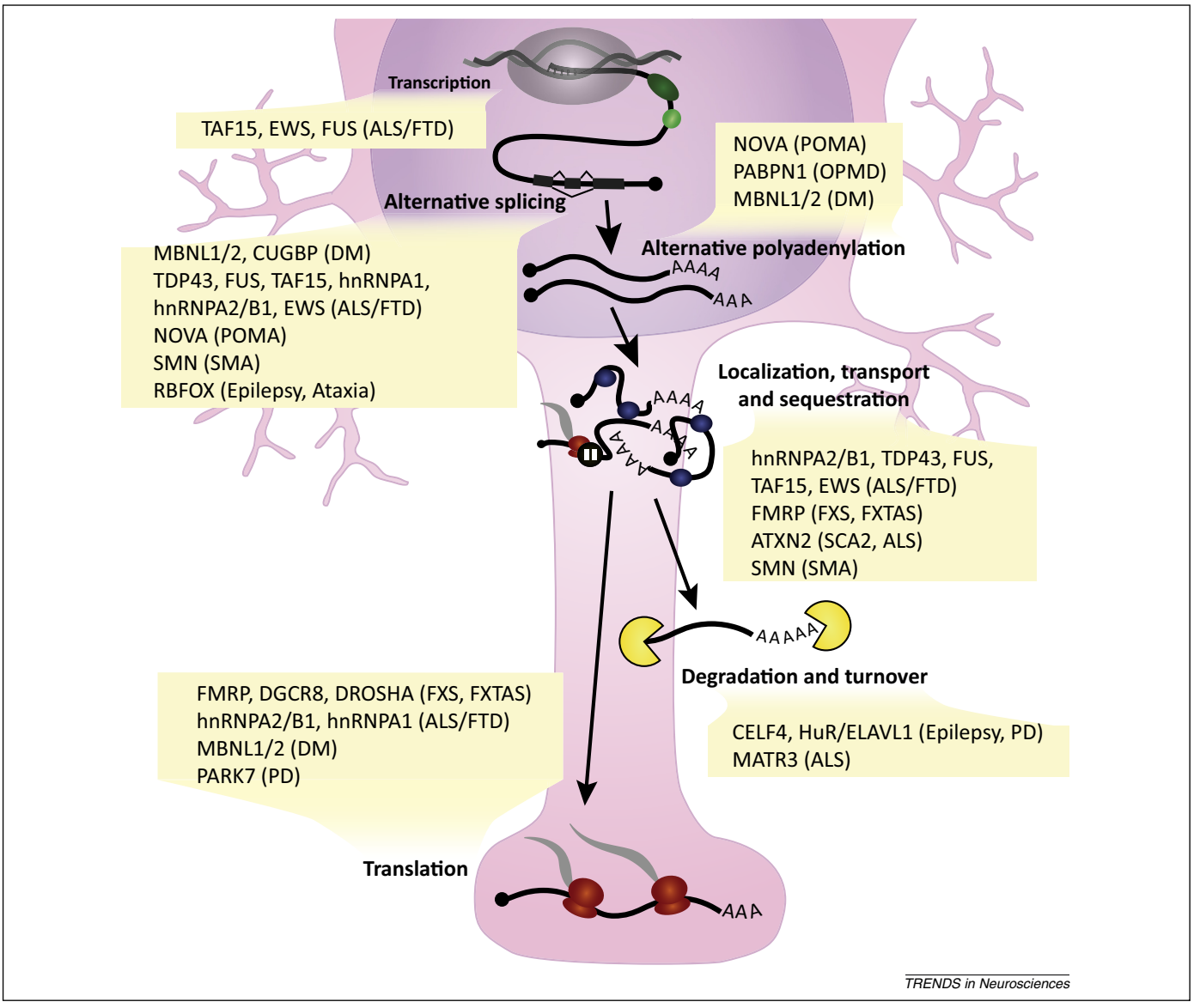

Figure 2. RNA-binding proteins (RBPs) are implicated in several neurological diseases. RBPs implicated in neurological diseases have a role in several steps of RNA processing, including transcription, alternative splicing, alternative polyadenylation, localization of transcripts, including sequestration into inclusions and stress granules, translation, RNA degradation, and turnover. The physiological (endogenous is also OK) functions of these RBPs often lend significant insights into the mechanisms of pathogenesis. Abbreviations: ALS, amyotrophic lateral sclerosis; DM, myotonic dystrophy; FTD, frontotemporal dementia; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor/ataxia syndrome; OPMD, oculopharyngeal muscular dystrophy; PD, Parkinson's disease; POMA, paraneoplastic opsoclonus-myoclonus ataxia; SCA2, spinocerebellar ataxia type 2; SMA, spinal muscular atrophy. For additional definitions, please see the main text.

mediated by coordinated roles of RBPs. Defects in polyadenylation and tail length undoubtedly result in aberrant gene expression and neuronal dysfunction; for example, dysfunction of polyadenylate-binding nuclear protein 1 (PABPN1), a protein involved in polymerization of the poly(A) tail, is implicated in oculopharyngeal muscular dystrophy (OPMD) (Figure 2).

Recently, sequencing-based methods have been developed to analyze alternative 3′ cleavage and polyadenylation (Figure 1D). Most of the early methods entail isolation of poly(A)+ RNA and oligo(d)T-primed cDNA synthesis. In PolyA-Seq [71], first strand synthesis (FSS) is oligo(d)T-primed and second strand synthesis is primed with a randomer. In poly(A) site sequencing (PAS-Seq) [72], FSS is carried out with a (d)T primer followed by ligation of adapters to both cDNA ends. Finally, in 3Seq [73], FSS is primed with an oligo(d)T containing a 3′ adapter. Internal priming is a major concern in these techniques, because sufficient sequence upstream of the poly(A) site must be identified to obtain mappable reads. An additional problem with (d)T priming is the propensity for polymerase slippage due to the repetitive nature of the poly(A) tail. To overcome this, a method called 3P-Seq [74] that experimentally defines 3′ ends independently of isolation of mRNA through poly(A)-stretches has also been used. 3P-Seq utilizes a splint ligation that attaches an adapter to the poly(A) tail as well as a poly(T)-adapter to the mRNA, thus avoiding poly(T) priming and simultaneously adding a biotin moiety to facilitate purification. Reverse transcription creates a stretch of (d)T complementary to the poly(A) tail that is partially digested with RNase H, leaving the sequence immediately adjacent to the site of polyadenylation available for sequencing. Notably, the quantitative ability of direct sequencing is not clear, because 3P-Seq

requires multiple enzymatic steps and adapter ligations, increasing chances of ligation-induced biases [75]. Another technique, termed '3′T-Fill' [76], incorporates a 'filling-in' step of the poly(A) tail with (d)T after libraries are clustered on the flow cell of the sequencing instrument, then starts sequencing at the site of polyadenylation.

Given that alternative polyadenylation is important for RNA localization, degradation, and translation, poly(A) sequencing has an important role in the study of neuronal function and neurological diseases. Variations on 3Seq and PAS-Seq were recently utilized to identify novel roles of PABPN1 in alternative polyadenylation. Loss of PABPN1 function in an OPMD mouse model, and cells expressing mutant trePABPN1, showed a widespread shift towards utilization of proximal polyadenylation sites resulting in a shorter 3′UTR [77,78]. Similarly, a role in polyadenylation was recently uncovered for Muscleblind-like (MBNL) proteins associated with DM [50]. Although it is not yet determined whether alternative splicing and alternative polyadenylation are coordinately regulated, there is increasing evidence that multiple levels of RNA processing are misregulated in neurological diseases.

In addition to alternative poly(A) site usage, poly(A) length is another relatively unexplored feature of mRNAs that may have a role in neurodegeneration, particularly in the case of PABPN1. Two techniques have recently been developed to determine poly(A) length: TAIL-Seq [79] and Poly(A)-tail length profiling by sequencing (PAL-Seq) [80] (Figure 1D). TAIL-Seq involves the ligation of a biotin-containing adapter to the 3′ end of mRNAs, followed by a partial digestion by RNase T1, which cleaves downstream of guanosine, thus protecting the poly(A) tail. The biotinylated RNA is streptavidin purified, 5′ phosphorylated, gel purified, ligated to a 5′ adapter, reverse-transcribed, amplified and paired-end sequenced. The first read uncovers the identity of the mRNA and the second determines the poly(A) length. However, one of the major obstacles to determining poly(A) length by sequencing is the residual fluorescent 'T' signal on the flow cell from sequencing long tracts of 'T's, which can drown out the signal of a non-T base and results in overestimating poly(A) tail length. To overcome this, TAIL-Seq incorporates an additional fluorescence reanalysis to determine the actual template base and, thus, significantly reduces overestimation of poly(A) length [79]. By contrast, PAL-Seq uses stochastic incorporation of a biotinylated uridine base that, when bound to fluorescently tagged streptavidin, gives a fluorescence intensity proportional to poly(A) length, although not with the same resolution as TAIL-Seq [80]. In this method, total RNA is 3′ splint ligated to a biotinylated adapter, partially digested, size selected, and biotin purified. RNAs are ligated to a 5′ adapter, reverse transcribed, and size selected. Next, sequencing clusters are generated on an Illumina flow cell, but before sequencing, a primer is hybridized 3′ to the terminal A of the poly(A) sequence and then extended into the poly(A) tail with dTTPs (deoxythymidine triphosphates) and biotin-conjugated dUTPs (deoxyuridine triphosphates). After sequencing into the poly(A) proximal region for mapping purposes, the flow cell is incubated with fluorophore-conjugated streptavidin, which binds the biotin and generates fluorescence proportional in intensity [based on a standard curve generated with synthetic poly(A) tails of known length] to poly(A) length. Performing TAIL-Seq and PAL-Seq on OPMD patient tissues will likely reveal differential poly(A) lengths in mRNA substrates, which may be key to uncovering the role of PABPN1 in disease pathogenesis.

Exploring translation

RBPs involved in translational control of mRNA have been associated with neurodegenerative diseases, such as FXS and fragile X-associated tremor/ataxia syndrome (FXTAS) (Figure 2). Sequencing and microarray analyses have been used to study translational control at the genome-wide level [7] (Figure 1E). A subset of these techniques utilizes sequencing of ribosome-associated mRNA populations, an approach termed 'polysome profiling', which uses the degree of ribosome occupancy of an mRNA as a measure of its translational efficiency. This approach entails isolation, purification, and sequencing of polysome-associated mRNA by sucrose gradient density centrifugation, and has been utilized to elucidate mechanisms of disease caused by FMRP in FXS [81]. Polysome profiling can be used for a variety of cell types, but does not allow isolation of cell type-specific polysomes in complex tissues, such as brain. Translating ribosome affinity purification (TRAP) utilizes a line of bacterial artificial chromosome (bacTRAP) transgenic mice with lineage-specific promoters driving expression of EGFP-tagged ribosomal protein L10 [82,83], enabling isolation of polysome-associated RNA from specific cell types by EGFP immunoaffinity purification. Similarly, Ribotag mice harbor a loxP-flanked version of the last exon of the ribosomal protein L22 gene (*Rpl22*) upstream of a hemagglutinin (HA)-tagged version of that same exon at the endogenous locus. Crossing Ribotag mice with lineage-specific Cre recombinase driver lines results in genomic deletion of the untagged exon and usage of the tagged exon, resulting in expression of HA-tagged L22 protein in the desired cell types [84]. Thus, this approach avoids a separate bacTRAP line for each cell type and uses the endogenous promoter to ensure near-physiological levels of the tagged protein.

A similar technique termed 'ribosome profiling' yields precise positions of translating ribosomes on mRNAs in a transcriptome-wide manner and, thus, results in deeper insights into gene-dependent dynamics of translational control [85,86]. In a manner reminiscent of CLIP, ribosome profiling involves treatment of cells with cycloheximide to prevent ribosome disassembly, followed by limited micrococcal nuclease-mediated digestion of RNA. The ribosome-protected mRNA fragments are isolated, sequenced, and mapped onto the genome. It will be informative to analyze the translational state in systems that allow recently discovered repeat-associated non-ATG translation (RAN) translation, which appears to be a common theme in microsatellite-associated neurological disorders [87,88].

Mechanisms and alterations of mRNA turnover in neurological diseases

The last phase of RNA processing is degradation, which can occur before or after the processing steps discussed above. RNA degradation is regulated by RBPs, some of which have been implicated in epilepsy and Parkinson's

disease, including CUGBP, Elav-like family member 4 (CELF4) [89] and Hu/ELAV [21], while others have been associated with ALS, including matrin 3 (MATR3) [90] (Figure 2). The current state-of-the-art in quantifying mRNA turnover rates is by metabolic labeling (Figure 1F). Cells are provided with synthetic nucleoside analogs containing a reactive functional group {typically 4-thiouridine (4sU) [91] or 5-ethynyl uridine (EU) [92]}, which are incorporated into nascent transcripts by the endogenous transcription machinery. Total RNA is then isolated and the labeled fraction biotinylated using label-specific chemistry. Using a pulse-chase technique, labeled RNA can be sequenced at various time points during the chase to quantify degradation rates of particular RNAs via RNA-Seq or qPCR. Labeled RNA can also be quantified during the pulse to determine synthesis rates. An analogous method, 5′-bromo-uridine (BrU) immunoprecipitation chase-deep sequencing analysis (BRIC-Seq) (Figure 1F), utilizes incorporation of BrU and immunoprecipitation of BrU-containing RNAs [93]. However, the benefit of the 4sU and EU methods is the strength of the streptavidin–biotin interaction, which enables stringent washing and cDNA synthesis on a solid support, eliminating an elution and precipitation step that can significantly reduce yields for downstream qPCR analyses. While CLIP has facilitated study of decay-associated RBPs involved in neurodegeneration, such as CELF4 [89] and ELAV [21], techniques that globally assay mRNA turnover could provide valuable insight into mechanisms of disease with pathological RNAs, such as those containing repeat expansions.

Secondary structure, post-transcriptional modifications, and editing of RNAs modulate RBP binding

Base-pairing, hairpins, bulges, multiloops, and unstructured regions in RNA have a significant role in the processing of transcripts, in part by mediating RBP binding and function. As mentioned above, current CLIP-Seq protocols do not readily identify structural preferences for RBP binding; however, several high-throughput methods have been developed to literally untangle the structure of RNA and the structural parameters of RBP targets (Figure 1G). In selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)-Seq [94], single-stranded loops and bulges are selectively acetylated, halting cDNA synthesis and generating sequencing libraries of various length corresponding to the location of single-stranded secondary structure at single-nucleotide resolution. To identify regions of secondary structure, fragmentation sequencing (FragSeq) utilizes a single-strand specific nuclease, leaving RNA with secondary structure for sequencing [95]. In addition, parallel analysis of RNA structure (PARS)-Seq compares sequencing libraries generated from treating mRNA with either a single-strand- or double-strand-specific nuclease to define secondary structure [96]. While these techniques address RNA secondary structure at a genome-wide level, they do not address RBP specificity for a particular structural motif. PIP-Seq may be able to address the RNA structure parameters necessary for RBP binding. In addition to technical advances, there are several computational tools for predicting structures within identified RBP binding sites. RBPmotif is a web server that utilizes such tools to predict secondary structure, including base-pairing, hairpin loops, bulge loops, multiloops, and unstructured regions [97]. No genome-wide studies have been done to examine the role of RNA structure in neurodegenerative disease; however, PIP-Seq and similar techniques could further elucidate the mechanism of pathology by identifying potentially altered structural requirements of RBP mutants in neurodegenerative disorders, or possibly a role of RNA structure in RBP sequestration and repeat expansion.

RNA also undergoes extensive post-transcriptional modifications, including methylation of the cap structure and the nucleobases of mRNA. Recent approaches have been developed to identify such marks on RNA (Figure 1G). For example, m6A-Seq [98], also called MeRIP-Seq [99], generates low-resolution genome-wide maps of N6-methyladenosine (m6A), a modification recently shown to have a role in transcript stability [100,101]. After RNA is fragmented, an antibody against m6A selects for m6A-containing fragments for sequencing. These techniques have not yet been directly applied to the study of neurodegenerative diseases, but given the recently identified role of m6A in transcript processing, doing so could reveal another layer of RNA-mediated pathology.

Finally, RNA is modified at the level of primary structure during a process called ‘RNA editing’, in which RNA deaminases convert adenosine to inosine (which is read as guanosine during reverse transcription) or cytosine to uracil [102]. Normal editing of glutamate receptor 2 (GluR2) was found to be significantly impaired in motor neurons from patients with ALS; this is a potential cause of excessive calcium influx that may contribute to motor neuron death [103–107]. In addition, abnormal editing of the glutamate transporter EAAT2 was identified in patients with ALS [108], and adenosine deaminase acting on RNA 2 (ADAR2) was recently proposed to bind expanded RNAs in patients with C9ORF72 ALS/FTD [109]. Finally, hnRNPA2/B1 recently linked to ALS/FTD [110] has been proposed as an enhancer of RNA editing [111], further emphasizing the potential role of editing in ALS/FTD. Single-end (SE) or paired-end (PE) RNA-Seq often proves sufficient to detect sequence changes due to RNA editing. So-called ‘padlock probes’ can be used to interrogate candidate-editing events by detecting single-base changes at tens of thousands of specific positions [112] (Figure 1G). The assay comprises an oligonucleotide designed such that its ends hybridize to isolated RNA to form a nearly circular structure at the target site except for a small gap. This gap is filled enzymatically and resulting sequences are compared between genomic DNA and cDNA, a base change being indicative of an editing event.

Concluding remarks

Neurological diseases are increasingly recognized as being associated with RNA regulatory dysfunction caused by reduced or aberrant RBP activity. Comprehensive analysis of the functions of these RBPs and identification of their target RNA regulatory networks are necessary to keep pace with the accelerated rate of their discovery in genetic diseases, and to enhance the development of therapeutics.

Box 1. Outstanding questions

- **What is the role of RNA secondary structure, editing, and modification in the pathology of neurodegenerative diseases?**
  Changes in RNA primary and secondary structure have yet to be thoroughly studied in the context of neurodegeneration. Of particular interest given the availability of new reagents and protocols are internal methylation and RNA secondary structure. With the recent evidence for a role of m6A in transcript stability, and potential roles in RNA localization and translation, applying m6A-Seq/MeRIP-Seq to a neurodegenerative model could identify misregulation of these processes through changes in RBP binding. Furthermore, for repeat expansion diseases, determining how secondary structure arising from long tracts of repeats influences RBP-mediated pathology is essential for understanding RBP sequestration and/or formation of foci.
- **How can we address the issue of heterogeneity in tissue and cellular models of neurodegeneration?**
  One of the greatest hurdles to understanding the mechanisms of RBP-mediated neuropathology is the complexity of brain tissue and populations of iPSC/ESC-derived neural cells. Advances in single cell isolation platforms, such as the Fluidigm C1 Auto-Prep System or other single-cell capture devices, are aiding in overcoming this hurdle, but there needs to be improvements in sequencing techniques to allow for a smaller amount of starting material. Techniques such as CLIP currently require millions of cells.
- **What will application of these techniques reveal about the mechanisms of neuropathology?**
  Few of the reviewed techniques aside from CLIP and its variations have been applied to the study of RBPs in the context of neuropathology, but doing so could reveal novel mechanisms of disease as well as potential therapeutic targets. As discussed here, several steps of RNA processing are misregulated in neurodegeneration, but the mechanisms remain elusive.
- **How can we study the RBPs and RNA transcripts found in foci and stress granules characteristic of several neurodegenerative diseases?**
  There is a growing need for a high-throughput technique for the study of foci and stress granules, and the associated RNA and RBPs, implicated in neurodegeneration. The inability to isolate these complexes has so far prevented such analyses, but once an efficient method has been developed for their isolation and purification, the associated RNA can be subjected to several of the reviewed techniques to better understand how stress granules form and the mechanism by which they contribute to disease pathology.

These analyses have been aided by the development of new genome-wide, high-throughput techniques to resolve the role of RBPs in RNA processing, as depicted in Figure 1.

Although genome-wide technologies have greatly enhanced our ability to study complex neurological diseases, several major challenges still exist (Box 1). One of the greatest hurdles to the study of disease, and cell biology in general, is the identification and isolation of rare cells that exhibit the phenotype of interest. Recent advances in single-cell technology are beginning to address these

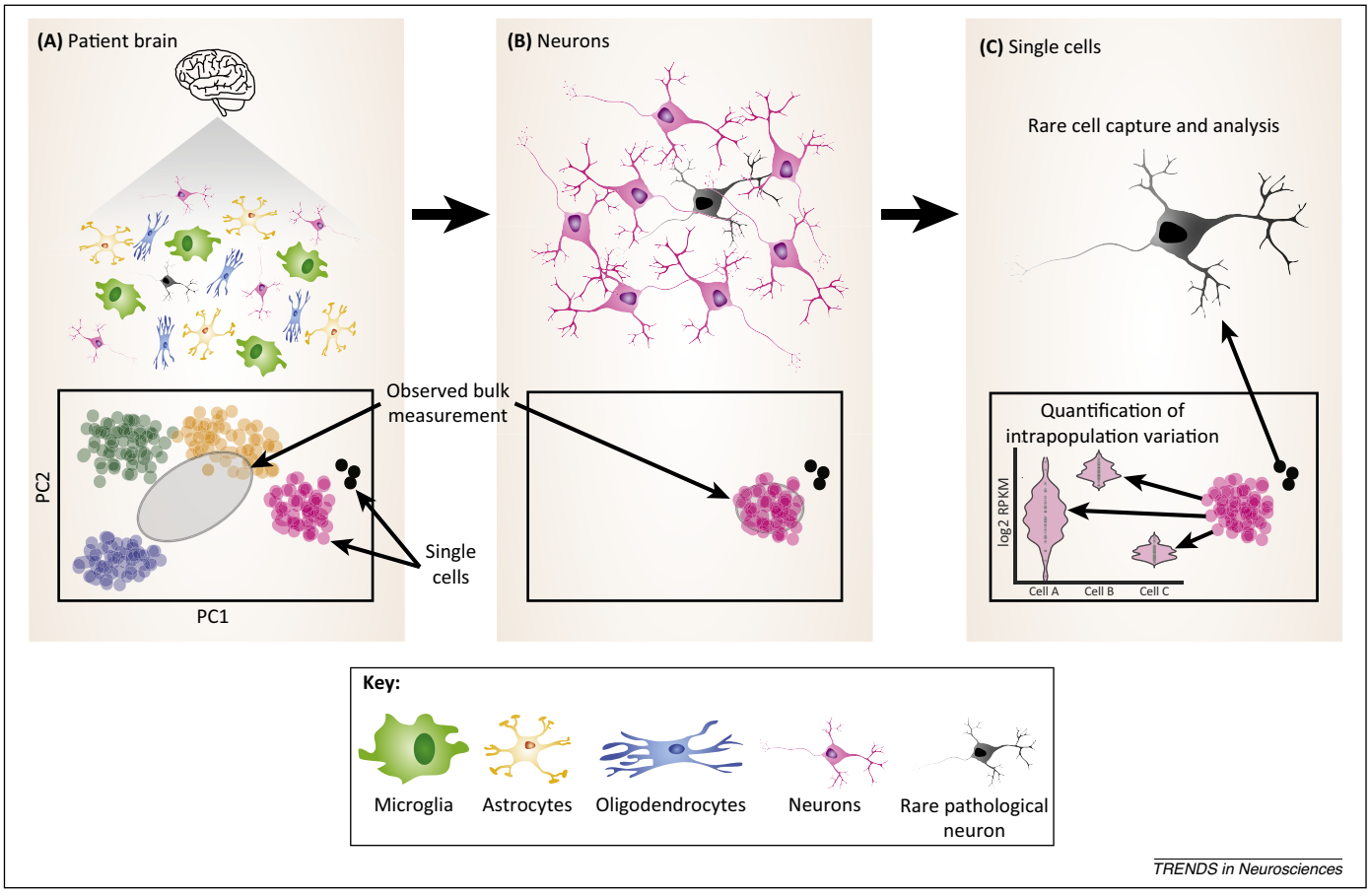

Figure 3. Single cell analysis identifies rare cell types and intrapopulation variation. **(A)** Analysis of bulk cell populations, such as brain tissue, prevents the identification of rare cell types, such as neurons with RNA –binding protein (RBP) inclusions, in addition to reducing evidence of phenotypic gradients by generating an average. **(B)** Isolation of a cell type may also fail to enable identification of a rare pathological cell. **(C)** With the isolation and analysis of single cell expression profiles, it becomes possible to identify not only rare cell types that differ from the bulk population, but also variation within a bulk population, such as differences in splicing.

deficits and resolve cellular heterogeneity (Figure 3). The ability to identify phenotypes with single cell resolution is particularly relevant for neurological studies, because the brain is naturally highly heterogeneous. Several neurological diseases, including DM, C9ORF72 ALS/FTD, and FXTAS, are characterized by nuclear RNA foci, which occur in only a subset of cells. Thus, when a bulk sample is studied, this small but critical subpopulation may be missed. In addition, repeat expansions can result in somatic mosaicism, and the gradient of phenotypes is lost in the average [113]. For example, repeat expansions as seen in ALS often differ in size between cell types, and this variability can account for variable pathology [114]. Furthermore, in diseases such as ALS/FTD that present with protein inclusions due to mutation or mislocalization of aggregation-prone RBPs, only a small percentage of cells have the pathological hallmarks. Application of single-cell technology is critical for overcoming issues of bulk sample heterogeneity and identifying such rare and variable cell types. There have been many developments in the single cell field, which are discussed in [115]. It would be ideal to combine single cell capture with the genome-wide techniques described here, but currently most of these methods require more material than can be obtained from single cells.

As the list of proteins and pathways linked to each neurodegenerative disease grows, directed approaches that focus on individual players are no longer sufficient to fully define the mechanisms of pathology. Rather, it is becoming necessary to adjust experimental approaches to identify characteristic global signatures of neurodegenerative diseases. To accomplish this, genome-wide approaches can identify comprehensive molecular snapshots, which can then be used to generate more accurate experimental models of disease, enhance diagnostic accuracy, and identify additional therapeutic targets. By using these techniques to observe the entire molecular 'forest' characteristic of the disease state, model systems can be developed to recapitulate the unique disease signature rather than modifying a single target that may not encompass the entire pathology. In addition to the novel and unique insights provided by genome-wide techniques, these methods also produce their share of challenges. Given that such a large number of targets and players are identified, there is the inherent problem of prioritizing which candidates to pursue. One aspect of this is determining the rate of false positives and false negatives, some of which can be addressed by utilizing appropriate controls. A second component of prioritization is to determine which targets are causative to the phenotype and which are merely bystanders, changing as the result of a change in a common upstream regulator. A third consideration is the emerging roles of noncoding (nc)RNAs that are often overlooked with current methodologies or computational analyses. Indeed, several (l)ncRNAs have been implicated in neurodegenerative disease, as reviewed in [116,117]. In addition to addressing this issue from a biological standpoint, there is also the issue of sheer computational power needed to wade through this ocean of data. This can be facilitated by the development of robust and efficient computational pipelines, but the ease with which such large data sets can be generated in today's age of benchtop sequencers requires further consideration to avoid a data-processing and analysis bottleneck. Finally, and perhaps most significantly, high-throughput genome-wide analyses are proving that RBPs have a significant role in the pathology of neurodegenerative diseases. As master regulators of several aspect of cellular function, it may prove necessary to target the availability of the normal versus mutant RBP, rather than its downstream targets and effectors. Some therapeutics are being developed to address this [109,118], including antisense oligonucleotides that directly block RBPs or indirectly block RBP binding by either altering RNA secondary structure or causing RNA degradation.

As these genome-wide techniques become more widely utilized, their application to the complex involvement of RBPs in neurological disease will bring much-needed clarity to understanding, modeling, and ultimately curing these devastating pathologies.

Acknowledgments

We are grateful to Dr. Don W. Cleveland and the members of his group, as well as the members of the Yeo lab for tremendous support and fruitful discussions. This work is supported by grants to G.W.Y. from the National Institute of Health (HG004659, NS075449, and U54HG007005) and California Institute for Regenerative Medicine (RB4-06045 and TR3-05676). G.W.Y. is an Alfred P. Sloan Research Fellow. R.B. is a Myotonic Dystrophy Foundation Fellow. C.L.T. is supported by grants from the National Institute of Health (R01NS087227 and P50AG005131), Target ALS (04827 and 0840), the Frick Foundation, and the Amyotrophic Lateral Sclerosis Association (grant 2235A). C.L.T. was a recipient of a Career Development Award from the Muscular Dystrophy Association and receives salary support from the Ludwig Institute for Cancer Research.

References

1 Lukong, K.E. *et al.* (2008) RNA-binding proteins in human genetic disease. *Trends Genet.* 24, 416–425
2 Cooper, T.A. *et al.* (2009) RNA and disease. *Cell* 136, 777–793
3 McKee, A.E. *et al.* (2005) A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev. Biol.* 5, 14
4 Yeo, G. *et al.* (2004) Variation in alternative splicing across human tissues. *Genome Biol.* 5, R74
5 Castle, J.C. *et al.* (2008) Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40, 1416–1425
6 Holt, C.E. and Bullock, S.L. (2009) Subcellular mRNA localization in animal cells and why it matters. *Science* 326, 1212–1216
7 Kapeli, K. and Yeo, G.W. (2012) Genome-wide approaches to dissect the roles of RNA binding proteins in translational control: implications for neurological diseases. *Front. Neurosci.* 6, 144
8 Darnell, J.C. and Klann, E. (2013) The translation of translational control by FMRP: therapeutic targets for FXS. *Nat. Neurosci.* 16, 1530–1536
9 Alami, N.H. *et al.* (2014) Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron* 81, 536–543
10 Baltz, A.G. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* 46, 674–690
11 Zhao, J. *et al.* (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* 40, 939–953
12 Mili, S. and Steitz, J.A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* 10, 1692–1694
13 Moore, M.J. *et al.* (2014) Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* 9, 263–293
14 Ule, J. *et al.* (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science* 302, 1212–1215

15 Licatalosi, D.D. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469
16 Darnell, R.B. (2013) RNA protein interaction in neurons. *Annu. Rev. Neurosci.* 36, 243–270
17 Wang, E.T. *et al.* (2012) Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* 150, 710–724
18 Nakaya, T. *et al.* (2013) FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* 19, 498–509
19 Ishigaki, S. *et al.* (2012) Position-dependent FUS–RNA interactions regulate alternative splicing events and transcriptions. *Sci. Rep.* 2, 529
20 Darnell, J.C. *et al.* (2011) FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261
21 Ince-Dunn, G. *et al.* (2012) Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* 75, 1067–1080
22 Charizanis, K. *et al.* (2012) Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron* 75, 437–450
23 Licatalosi, D.D. *et al.* (2012) Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes Dev.* 26, 1626–1642
24 Weyn-Vanhentenryck, S.M. *et al.* (2014) HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152
25 Yeo, G.W. *et al.* (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 16, 130–137
26 Lagier-Tourenne, C. *et al.* (2012) Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* 15, 1488–1497
27 Polymenidou, M. *et al.* (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* 14, 459–468
28 Sanford, J.R. *et al.* (2009) Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* 19, 381–394
29 Modic, M. *et al.* (2013) CLIPing the brain: studies of protein-RNA interactions important for neurodegenerative disorders. *Mol. Cell. Neurosci.* 56, 429–435
30 Hafner, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141
31 Kishore, S. *et al.* (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat. Methods* 8, 559–564
32 Burger, K. *et al.* (2013) 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol.* 10, 1623–1630
33 Konig, J. *et al.* (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915
34 Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346, 818–822
35 Campbell, Z.T. *et al.* (2012) Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Rep.* 1, 570–581
36 Ray, D. *et al.* (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* 27, 667–670
37 Ray, D. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177
38 Lambert, N. *et al.* (2014) RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol. Cell* 54, 887–900
39 Castello, A. *et al.* (2013) System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* 8, 491–500
40 Meisenheimer, K.M. and Koch, T.H. (1997) Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* 32, 101–140
41 Silverman, I.M. *et al.* (2014) RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol.* 15, R3
42 Hoell, J.I. *et al.* (2011) RNA targets of wild-type and mutant FET family proteins. *Nat. Struct. Mol. Biol.* 18, 1428–1431
43 Wilbert, M.L. *et al.* (2012) LIN28 binds messenger RNAs at GGAGA motifs and regulates splicing factor abundance. *Mol. Cell* 48, 195–206
44 Cho, J. *et al.* (2012) LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* 151, 765–777
45 Sephton, C.F. *et al.* (2011) Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* 286, 1204–1215
46 Xiao, S. *et al.* (2011) RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell. Neurosci.* 47, 167–180
47 Colombrita, C. *et al.* (2012) TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *J. Biol. Chem.* 287, 15635–15647
48 Rogelj, B. *et al.* (2012) Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* 2, 603
49 Tollervey, J.R. *et al.* (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* 14, 452–458
50 Batra, R. *et al.* (2014) Loss of MBNL Leads to disruption of developmentally regulated alternative polyadenylation in RNA-mediated disease. *Mol. Cell* 56, 311–322
51 van der Brug, M.P. *et al.* (2008) RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10244–10249
52 Blackinton, J. *et al.* (2009) Post-transcriptional regulation of mRNA associated with DJ-1 in sporadic Parkinson disease. *Neurosci. Lett.* 452, 8–11
53 Pan, Q. *et al.* (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415
54 Wang, Z. and Burge, C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14, 802–813
55 Mills, J.D. and Janitz, M. (2012) Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases. *Neurobiol. Aging* 33, 1012.e11–1012.e24
56 Johnson, J.M. *et al.* (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144
57 Pan, Q. *et al.* (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* 16, 929–941
58 Katz, Y. *et al.* (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015
59 Li, H. *et al.* (2008) Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20179–20184
60 Lovci, M.T. *et al.* (2013) Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* 20, 1434–1442
61 Zhou, Z. *et al.* (2012) The Akt-SRPK-SR axis constitutes a major pathway in transducing EGF signaling to regulate alternative splicing in the nucleus. *Mol. Cell* 47, 422–433
62 Li, H. *et al.* (2012) RASL-seq for massively parallel and quantitative analysis of gene expression. *Curr. Protoc. Mol. Biol.* 4, 1–9
63 Wuarin, J. and Schibler, U. (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* 14, 7219–7225
64 Wetterberg, I. *et al.* (2001) In situ transcription and splicing in the Balbiani ring 3 gene. *EMBO J.* 20, 2564–2574
65 Zhou, Y. *et al.* (2013) ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. *PLoS Genet.* 9, e1003895
66 Bertolotti, A. *et al.* (1998) EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Mol. Cell. Biol.* 18, 1489–1497
67 Schwartz, J.C. *et al.* (2012) FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* 26, 2690–2695

68 Khodor, Y.L. *et al.* (2011) Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* 25, 2502–2512
69 Core, L.J. *et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848
70 Churchman, L.S. and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373
71 Derti, A. *et al.* (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183
72 Shepard, P.J. *et al.* (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* 17, 761–772
73 Beck, A.H. *et al.* (2010) 3′-end sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS ONE* 5, e8768
74 Jan, C.H. *et al.* (2011) Formation, regulation and evolution of *Caenorhabditis elegans* 3′UTRs. *Nature* 469, 97–101
75 Hafner, M. *et al.* (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* 17, 1697–1712
76 Wilkening, S. *et al.* (2013) An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* 41, e65
77 Jenal, M. *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149, 538–553
78 de Klerk, E. *et al.* (2012) Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* 40, 9089–9101
79 Chang, H. *et al.* (2014) TAIL-seq: Genome-wide determination of Poly(A) tail length and 3′ end modifications. *Mol. Cell* 53, 1044–1052
80 Subtelny, A.O. *et al.* (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71
81 Stefani, G. *et al.* (2004) Fragile X mental retardation protein is associated with translating polyribosomes in neuronal cells. *J. Neurosci.* 24, 7272–7276
82 Heiman, M. *et al.* (2008) A translational profiling approach for the molecular characterization of CNS cell types. *Cell* 135, 738–748
83 Doyle, J.P. *et al.* (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135, 749–762
84 Sanz, E. *et al.* (2009) Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13939–13944
85 Ingolia, N.T. *et al.* (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223
86 Ingolia, N.T. *et al.* (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* 7, 1534–1550
87 Zu, T. *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 260–265
88 Cleary, J.D. and Ranum, L.P. (2013) Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum. Mol. Genet.* 22 (R1), R45–R51
89 Wagnon, J.L. *et al.* (2012) CELF4 regulates translation and local abundance of a vast set of mRNAs, including genes associated with regulation of synaptic function. *PLoS Genet.* 8, e1003067
90 Johnson, J.O. *et al.* (2014) Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat. Neurosci.* 17, 664–666
91 Rabani, M. *et al.* (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* 29, 436–442
92 Jao, C.Y. and Salic, A. (2008) Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15779–15784
93 Imamachi, N. *et al.* (2013) BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. *Methods* 67, 55–63
94 Mortimer, S.A. *et al.* (2012) SHAPE-Seq: high-throughput RNA structure analysis. *Curr. Protoc. Chem. Biol.* 4, 275–297
95 Underwood, J.G. *et al.* (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001
96 Kertesz, M. *et al.* (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107
97 Kazan, H. and Morris, Q. (2013) RBPmotif: a web server for the discovery of sequence and structure preferences of RNA-binding proteins. *Nucleic Acids Res.* 41, W180–W186
98 Dominissini, D. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206
99 Meyer, K.D. *et al.* (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* 149, 1635–1646
100 Wang, X. *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120
101 Wang, Y. *et al.* (2014) N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat. Cell Biol.* 16, 191–198
102 Farajollahi, S. and Maas, S. (2010) Molecular diversity through RNA editing: a balancing act. *Trends Genet.* 26, 221–230
103 Hideyama, T. *et al.* (2010) Induced loss of ADAR2 engenders slow death of motor neurons from Q/R site-unedited GluR2. *J. Neurosci.* 30, 11917–11925
104 Kawahara, Y. *et al.* (2004) Glutamate receptors: RNA editing and death of motor neurons. *Nature* 427, 801
105 Kwak, S. and Kawahara, Y. (2005) Deficient RNA editing of GluR2 and neuronal death in amyotropic lateral sclerosis. *J. Mol. Med.* 83, 110–120
106 Aizawa, H. *et al.* (2010) TDP-43 pathology in sporadic ALS occurs in motor neurons lacking the RNA editing enzyme ADAR2. *Acta Neuropathol.* 120, 75–84
107 Hideyama, T. *et al.* (2012) Profound downregulation of the RNA editing enzyme ADAR2 in ALS spinal motor neurons. *Neurobiol. Dis.* 45, 1121–1128
108 Flomen, R. and Makoff, A. (2011) Increased RNA editing in EAAT2 pre-mRNA from amyotrophic lateral sclerosis patients: involvement of a cryptic polyadenylation site. *Neurosci. Lett.* 497, 139–143
109 Donnelly, C.J. *et al.* (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* 80, 415–428
110 Kim, H.J. *et al.* (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 495, 467–473
111 Garncarz, W. *et al.* (2013) A high-throughput screen to identify enhancers of ADAR-mediated RNA-editing. *RNA Biol.* 10, 192–204
112 Li, J.B. *et al.* (2009) Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 324, 1210–1213
113 McConnell, M.J. *et al.* (2013) Mosaic copy number variation in human neurons. *Science* 342, 632–637
114 Dols-Icardo, O. *et al.* (2014) Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. *Hum. Mol. Genet.* 23, 749–754
115 Shapiro, E. *et al.* (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630
116 Wu, P. *et al.* (2013) Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain Res. Bull.* 97, 69–80
117 Salta, E. and De Strooper, B. (2012) Non-coding RNAs with essential roles in neurodegenerative disorders. *Lancet Neurol.* 11, 189–200
118 Wheeler, T.M. *et al.* (2012) Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature* 488, 111–115
119 Elkon, R. *et al.* (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.* 14, 496–506
120 Balagopal, V. *et al.* (2012) Ways and means of eukaryotic mRNA decay. *Biochim. Biophys. Acta* 1819, 593–603
121 He, F. and Jacobson, A. (2001) Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. *Mol. Cell. Biol.* 21, 1515–1530
122 Frischmeyer, P.A. *et al.* (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* 295, 2258–2261
123 van Hoof, A. *et al.* (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science* 295, 2262–2264
124 Zhang, C. and Darnell, R.B. (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614
125 Castello, A. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406
126 Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476