RNA-binding proteins in human genetic disease

Fátima Gebauer[1,2,5†], Thomas Schwarzl[3,5], Juan Valcárcel[1,2,4] and Matthias W. Hentze[3†]

[1]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, 08003-Barcelona, Spain
[2]University Pompeu Fabra (UPF), 08003-Barcelona, Spain
[3]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg, Germany.
[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
[5]These authors contributed equally: Fátima Gebauer, Thomas Schwarzl
[†]e-mail: hentze@embl.org, fatima.gebauer@crg.eu

https://doi.org/10.1038/s41576-020-00302-y

Abstract

RNA-binding proteins (RBPs) are critical effectors of gene expression, and as such their malfunction underlies the origin of many diseases. RBPs can recognize hundreds of transcripts and form extensive regulatory networks that help to maintain cell homeostasis. System-wide unbiased identification of RBPs has increased the number of recognized RBPs into the four-digit range and revealed new paradigms: from the prevalence of structurally disordered RNA-binding regions with roles in the formation of membraneless organelles, to unsuspected and potentially pervasive connections between intermediary metabolism and RNA regulation. Together with an increasingly detailed understanding of molecular mechanisms of RBP function, these insights are facilitating the development of new therapies to treat malignancies. Here, we provide an overview of RBPs involved in human genetic disorders, both Mendelian and somatic, and discuss emerging aspects in the field with emphasis on molecular mechanisms of disease and therapeutic interventions.

Introduction

RNA-binding proteins (RBPs) assemble with RNA to form ribonucleoprotein particles (RNPs) that are dynamic in nature, as RNP composition changes according to the maturation or functional state of RNA as well as the cellular context[1-3]. RBPs regulate all aspects of RNA life, including transcription, splicing, modification, intracellular trafficking, translation and decay (Figure 1). Conversely, RNA may regulate the activity or location of RBPs, a feature that has been termed "riboregulation"[4]. A prime example of riboregulation is protein kinase R (PKR), the binding of which to double-stranded RNA induces protein dimerization and autophosphorylation, resulting in activation of the enzyme[5].

Classic RBPs are characterized by the presence of one or more RNA-binding domains (RBDs). Most RBDs show defined 3D structures or features that make them

computationally predictable. Classic RBDs include the prevalent RNA recognition motif (RRM), the RGG box, the K-homology (KH), DEAD/DEAH helicase and zinc-finger domains, and around 30 other domains of lesser abundance[2]. Recent unbiased RNA interactome approaches (Figure 2) have revealed additional unconventional RBPs that lack discernible RBDs but frequently contain intrinsically disordered regions (IDRs) or mononucleotide and dinucleotide binding domains that directly engage in RNA binding[4,6-8]. Because the known cell biological functions of these proteins are unrelated to RNA biology, they have been coined "enigmRBPs"[4,9].

Compared with transcription factors, RBPs are more evolutionarily conserved and more widely distributed across tissues, consistent with their frequent housekeeping roles[2]. Despite these properties, mutations or alterations in the expression of housekeeping RBPs tend to cause tissue-specific defects. How is this plausible? First, RBPs may act on RNA targets or with regulatory partners that display tissue-specific expression. Second, RBPs can bind RNA targets with a wide range of affinities and specificities, modulated by post-translational modifications, interacting partners and local sequence or structure context in the RNA, leading to the formation of dynamic and cell type-specific regulatory complexes[10-13]. Third, RNA binding *per se* does not necessarily elicit regulatory effects. Although RBPs can bind hundreds of RNA targets, only subsets may be regulated under particular cellular conditions. Groups of RNAs coordinately regulated by an RBP under a given stimulus are dubbed RNA regulons[14,15]. Last, the extensive networks that RBPs form with their RNA targets and other proteins are characterized by redundancy, feedback and feedforward control, which together provide robustness such that alterations may be differentially buffered in one cell type versus another.

The basic properties and functions of RBPs have been expertly discussed in recent reviews, including those mentioned above. Here, we focus on the effect of RBP mutations in human genetic disorders. We provide an updated overview of RBPs mutated in genetic disease and highlight the emerging relevance of disordered regions. We then select concrete examples to illustrate underlying molecular mechanisms of disease, and discuss current therapeutic avenues. Other reviews on one or several aspects of these subjects can be found elsewhere[16-19].

RNA binding proteins in genetic disease

The RBP family is one of the largest protein groups in the cell. Earlier RNA interactome capture (RIC) studies have generated an RBP superset of 1,393 members that bind to polyadenylated RNA in human cells[4]. To provide a comprehensive overview of RBPs mutated in genetic disease, we have updated the RBP list, and included results from unbiased studies for total RNA, which increases the integrated number of candidate RBPs from different human cell types to 4,257. These studies include RIC[6-9,20] as well as the related methods enhanced RIC (eRIC)[21,22], RBDmap[23], pCLAP (peptide crosslinking and affinity purification)[24] and serial interactome capture (serIC)[25], and various other

methods that select RBPs crosslinked to RNA irrespective of its poly(A) status, including CARIC (chemistry-assisted RNA interactome capture)[26], RICK (RNA interactome using click chemistry)[27], OOPS (orthogonal organic phase separation)[28], XRNAX[29], PTex (phenol-toluol extraction)[30] and RNPxl[31] (Figure 2 and Table S1). Because these methods all employ UV-crosslinking and mass spectrometry, proteins of low abundance or those that do not crosslink well to RNA can be missed. Hence, the number of RBPs may further increase in the future. These updated data are available in RBPbase (https://rbpbase.shiny.embl.de/) and Table S1.

Previously annotated RBPs are consistently detected in RIC studies, whereas a large number of additional candidate RBPs have also been reported (Figure 3a). It is important to note that UV-crosslinking based approaches are not equilibrium methods and, consequently, do not assess RBP affinity or specificity. UV-crosslinking shows technical biases (for example, uridines crosslink better than other bases), and even transient interactions, given the correct geometry, may be detected efficiently[32]. RBPs detected by RIC should therefore be confirmed using orthogonal methods, such as the polynucleotide kinase (PNK) assay or gel-mobility shift assays to confirm RNA binding and assess affinity and specificity parameters. CLIP (crosslinking and immunoprecipitation) assays to identify RNA targets and binding signatures, and mutational analysis to identify RBDs are commonly used steps to investigate the function of RBP-RNA interactions. Although RICs provide a powerful tool to detect biochemical RNA binding in cells, only functional follow-up experiments can establish the biological relevance of newly detected RBPs.

For our analyses, we considered a superset of 3,470 RBPs consisting of 2,650 RBPs that have been detected in at least two RIC studies combined with the previously annotated RBPs (Figure 3b). Detection in at least two independent RIC experiments reduces the probability of false positive assignments, especially when those methods follow different biochemical principles (see Figure 2). Only 25% of the RBPs in the RBP superset harbour RBDs that can be predicted bioinformatically (Figure 3c).

To identify RBPs mutated in genetic disease, we crossed our RBP superset with disease association data from the Open Targets platform (https://www.opentargets.org/). We consider genetic disease broadly, including both Mendelian and somatic mutations. Nearly one-third of the superset (1,054 RBPs) were mutated in disease, accounting for >20% of all proteins (4,912) with annotated mutations (Figure 3d). In comparison, transcription factors account for 10% of all proteins with annotated mutations, with a proportionally lesser contribution to Mendelian mutations (Figure 3e). We provide a summary in Table S2. Notably, Gene Ontology (GO) analysis shows that mutated RBPs are predominantly associated with metabolism and nervous system development, highlighting connections between metabolism, RNA regulation and disease, conceivably with increased prevalence in neurological pathologies (Figure 3f). Indeed, diseases of the nervous system are the top Mendelian disorders affected by mutations in RBPs (Figure 3g), with a clear prevalence

of RBPs lacking classic RBDs (Figure 3h). The incidence of mutations in RBPs containing or lacking classical RBDs varies with the type of disease (Figure 3h). Strikingly, RPBs outnumber transcription factors and other proteins regarding the prevalence of mutations in numerous Mendelian disorders. However, this picture is inverted when one considers somatic mutations, where transcription factors take the lead (Figure 3g,h). A dramatic case of this trend is observed in cell proliferation disorders, a category including cancer, consistent with previous reports showing that somatic mutations in transcription factors are more frequently detected in cancerous cells[33]. As inherited mutations must be compatible with embryonic and fetal development, one could speculate that intrauterine development may be more permissive to mutations in RBPs, which display housekeeping functions that are supported by buffering systems.

RBP mutations and mechanisms of disease

There are many ways in which a somatic mutation or a germline mutation may influence the function of an RBP (Figure 4). First, mutations in genes encoding RBPs can alter their expression levels, including the relative ratio between alternative isoforms that display distinct interactions and activities, or indirectly influence RBP function (for example, see FMRP discussed below). Second, mutations may truncate the protein or change its amino acid composition such that interactions with cofactors, RNA targets or metabolites are altered. For instance, the protein HuR (also known as ELAV-like protein 1) binds to the metabolite UDP-glucose, which prevents HuR association with SNAI1 mRNA. Mutations in HuR that abrogate UDP-glucose binding occur in various cancer cells, and result in increased binding and stabilization of SNAI1 mRNA, which encodes an epithelial to mesenchymal transition marker, leading to a cellular gain of invasive properties[34]. Third, mutations may affect the (enzymatic) properties of RBPs that have dual roles as enzymes. For example, mutations in the translation initiation factor complex eIF2B cause leukoencephalopathy with vanishing white matter (OMIM #603896), a fatal widespread loss of brain glial content. These mutations usually reduce the guanine-nucleotide exchange factor (GEF) activity of this complex, leading to reduced levels of eIF2-GTP and to an overall decrease of translation[35]. Fourth, mutations can cause mislocalization and/or aggregation of the protein (for example, see the FUS (fused in sarcoma) discussed below). Last, mutations in RBP binding sites can influence their regulation *by* RBPs. For example, mutations in the iron-responsive element of the gene FTL affect IRP1 binding to FTL mRNA and result in hyperferritinaemia-cataract syndrome (OMIM #600886)[36].

Mutations in RBPs are found more often in functional domains outside RBDs (Figure 5a). Intriguingly, somatic mutations also frequently affect the untranslated regions (UTRs) of transcripts encoding RBPs compared with other proteins or transcription factors (Figure 5a). This finding points to de-regulation of RBP expression at the post-transcriptional level in disease, and to cross-regulatory RBP networks that would be interesting to decipher in future studies. Protein–protein interaction networks

using STRING[37] indeed reveal extensive connections among RBPs. Two examples are shown in Figures 5b-c for Alzheimer disease and dyskeratosis congenita, respectively, where contacts with and among RBPs are extensive and central, involving founding disease members such as amyloid precursor protein (APP) and dyskerin (DKC1) as RBPs.

We next illustrate the variety of disease mechanisms caused by RBP mutations with select examples that show the relevance of RBP dynamics in the generation of disease, and the surprising phenotypic specificity of mutations in RBPs that perform general functions.

Disordered regions, phase transitions and neurodegenerative disorders

Some RBPs display the propensity to coalesce into membraneless compartments by a process known as liquid-liquid phase separation, where a homogeneous solution of macromolecules de-mixes into a dense phase rich in macromolecules and a surrounding dilute phase. The assembly of membraneless compartments involves the establishment of multiple weak interactions between RNA molecules and RBPs, generally engaging low-complexity, disordered regions of RBPs (and other proteins)[38-41]. The nature of this network of interactions allows the exchange of macromolecules with the surrounding cellular environment and makes these compartments dynamic and reversible. It has been proposed that cell compartments such as the nucleolus, and a suite of RNP granules including P-bodies, stress granules, P-granules in *Caenorhabditis elegans*, paraspeckles and Cajal bodies, among others, are formed following the principles of phase separation[42-44] (reviewed in ref. 45). However, the involvement of true liquid–liquid phase separation in the formation of these compartments has been challenged[46].

Condensates assembled by phase separation can transit different material states. Although initial de-mixing generally yields macromolecular condensates with liquid-like properties, these assemblies can progress into hydrogel-like states with reduced fluidity, and even into dense pathological aggregates (Figure 6a). Regarding function, liquid-like compartments may foster concentrated biochemical reactions, whereas more viscous assemblies could serve to inactivate or store macromolecules. However, a clear causal link between cellular condensates, material properties and function is still missing. Pathological protein aggregates, which are common in neurodegenerative disorders[47,48], are solid-like inclusions that are irreversible. Indeed, many degenerative brain diseases are characterized by the deposition of toxic protein aggregates containing RBPs, for example, of amyloid-β and Tau in Alzheimer disease, α-synuclein in Parkinson disease, and FUS, EWS and TAF15 (also referred to as FET proteins) or TDP43 in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[40]. Numerous disease-associated RBPs have disordered prion-like domains, which promote protein-protein interactions that can accelerate liquid to solid phase transitions[40,41,47,48].

Recent advances in phase separation research are exemplified by the protein FUS (Fused in sarcoma). FUS is a multifunctional RBP involved in practically all steps of

gene expression, although it is best known as a transcription, DNA damage repair and splicing factor. FUS contains two types of low-complexity regions that synergize in phase separation, an amino-terminal prion-like domain rich in QGSY residues; and carboxy-terminal RGG-rich regions that contribute to RNA binding[38,48] (Figure 6b). Interactions of Y residues of the former domain with R residues of the latter primarily govern phase separation, whereas Q, G and S residues modulate fluidity, revealing a molecular grammar with potential predictive capacity[49]. Mutations in some of these residues accelerate the formation of amyloid aggregates and correlate with neuronal toxicity in fly models, linking amyloid formation with neurodegeneration[47-49]. Importantly, RGG regions are in close proximity to nuclear localization signals, and a large fraction of FUS mutations cause its cytoplasmic localization[47,50,51]. The relative contribution of loss of nuclear function versus toxic aggregation in the cytoplasm to disease is unclear.

The dynamics of phase separation may change with age. It has been speculated that gradual loss of gene expression control is accompanied by changes in the concentration and heterogeneity of macromolecules, accumulation of misfolded proteins and decline in mitochondrial activity, circumstances that promote the appearance of amyloid aggregates[52]. Together with defective clearance of these aggregates owing to a decline in the activity of the proteasome and autophagy machineries, these factors could potentially explain why the appearance of amyloid-related neurodegenerative disorders increases with age in humans.

Fragile X: a paradigm of nucleotide repeat expansion disorders and their complexity

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability, and the first described example of a nucleotide repeat expansion disorder[53]. It is caused by CGG triplet expansions located within the 5' UTR of the *FMR1* gene, the number of excessive repeats correlating with the severity (and type) of disease: genotypes of 5–44 repeats have no phenotype; those between 55 and 200 repeats (also called premutation repeats) cause fragile X-associated tremor ataxia syndrome (FXTAS); and expansions >230 repeats lead to hypermethylation and silencing of the *FMR1* gene, resulting in FXS[54].

The *FMR1* gene encodes FMRP, an RBP involved in transport and translation of mRNAs in neurons[54-57]. FMRP contains two KH RBDs and one RGG region (Figure 7a). The relevance of FMRP RNA binding in disease was supported by an early report of a patient with FXS with a normal triplet repeat number but bearing a point mutation (Ile304Asn) in the KH2 RBD[58,59]. It was later found that FMRP associates with polysomes, and the Ile304Asn mutation abrogates this association, suggesting roles in translational control[60]. Efforts to precisely understand FMRP RNA-binding specificity and function are ongoing, focusing on mechanisms by which this protein can control translation. In vitro RNA selection experiments showed that the RGG box of FMRP binds to specific G-rich sequences, so-called G-quartets[61], whereas the KH2 domain recognizes an intricate tertiary structure referred to as the 'kissing complex'[62]. However, studies performed in

vivo failed to detect these elements within FMRP target RNAs, yet found a predominant association of FMRP with open reading frames and a role in stalling elongating ribosomes[57]. Such a role is supported by recent cryogenic electron microscopy structural data showing that the KH domains of FMRP interact directly with the ribosome near the P-site, whereas the RGG box is free to potentially interact with mRNA targets[63]. Additional reports have argued that FMRP binds to mRNA via short sequence elements[64], structured motifs[65] or indirectly via non-coding RNA[66]. Similar to the apparent diversity of binding modes, alternative mechanisms of translational control have been put forward, such as repression of translation initiation[56], modulation of microRNA-mediated repression[67-69], and even translation activation[65,70]. Although these scenarios are not mutually exclusive, further work is required to resolve the roles of FMRP in translational control.

The fact that different numbers of CGG repeats in the 5' UTR of the *FMR1* gene elicit different syndromes is intriguing. Contrary to FXS, FXTAS is characterized by a significant increase in *FMR1* mRNA with only slight reductions in FMRP levels, which originally suggested an RNA-mediated mechanism in the pathogenesis of the disease. A hallmark of FXTAS is the accumulation of ubiquitin-positive intranuclear inclusions containing *FMR1* mRNA[71] (Figure 7b). These inclusions are detected throughout the central nervous system of patients with FXTAS in a manner that correlates with the number of CGG repeats, and are absent in patients with FXS (who lack expression of *FMR1*), suggesting a role in the disease[72]. Research using FXTAS fly models or mammalian cells implicated premutation repeats in the formation of nuclear RNA inclusions that function as 'RBP sinks', sequestering the trapped RBPs from their normal functions[73-76]. This RNA-driven mechanism of FXTAS has not been confirmed in mouse models; in mice, no RNA inclusions formed by premutation repeats alone were detected, and data point to unconventional repeat-associated non-AUG initiated (RAN) translation as the primary mechanism of disease. RAN translation leads to production of toxic polyglycine-containing FMRP peptides that can be detected in inclusions[77] and seem to be responsible for neurodegeneration, potentially by interacting with the nuclear lamina[78].

FMR1 is ubiquitously expressed, with higher expression in the brain, thyroid gland, and female reproductive organs. Given the important functions of FMRP, it is not surprising that its complete absence has dramatic consequences, especially for the brain and gonadal tissues. What is less clear is why other tissues are not phenotypically affected. Perhaps the complexity inherent to the brain and gonads has evolved at the expense of an equally complex post-transcriptional regulatory programme, creating strong dependencies on RBPs such as FMRP.

Mutations in gene-specific and general RNA processing factors: RBM10 and PRP8

Mutations in the gene encoding the RBP RBM10 underlie the X-linked pleiotropic developmental TARP syndrome (OMIM #311900)[79,80]. RBM10 can function as a splicing

regulator and, indeed, alterations in alternative splicing have been identified in lymphoblastoid cell lines derived from a patient with TARP syndrome, as well as in mandibular cells from *Rbm10* knockout mice and mouse embryonic stem cells depleted of RBM10 (refs 81, 82). These findings suggest that splicing alterations are mechanistically involved in the pathogenesis of TARP syndrome. Possibly related to heart abnormalities in this syndrome, RBM10 can also regulate mRNA 3'-end formation of genes involved in cardiac development[83].

In addition to TARP syndrome, somatic mutations in the *RBM10* gene occur in numerous cancers, including lung adenocarcinoma[84-87], colorectal cancer[88], pancreatic cancer[89] and bladder cancer[90]. Consistent with context-specific functions of RBPs in cancer, RBM10 can function as a tumour suppressor or an oncogene[91-93], possibly linked to expression of alternatively spliced variants and cross-regulatory feedback loops with the related protein RBM5 (ref. 94). The functions of RBM10 as a tumour suppressor can, at least in part, be explained by modulation of alternative splicing of genes relevant for the control of cell growth and apoptosis, including the Notch regulator NUMB[91,95,96]. Mechanistically, RBM10 binds to specific sequence motifs in the pre-mRNA and inhibits inclusion of exons near RBM10 binding sites[81,91,97], leading to defects on specific genes and to precise outcomes in the context of development and cancer. In contrast to RBM10, mutations in conserved core splicing factors -in principle, required for removal of every intron- would be expected to have an impact on general splicing and widespread cellular functions and cell viability. Remarkably, however, some such mutants lead to highly specific developmental defects. For example, PRP8 (also known as PRPF8) is a central component of spliceosomes that chaperones the RNA-based catalytic centre, is in close contact with the splice sites and, through protein–protein interactions, modulates the activity of BRR2, a helicase that initiates key RNA–RNA rearrangements leading to splicing catalysis[98-101]. Mutations in the C-terminal domain of PRP8 disrupt the interaction with BRR2 and cause a severe form of the eye condition retinitis pigmentosa[102-104]. Rather than overall defects in RNA processing, disrupted splicing of the autophagy regulatory gene *ULK1* leads to impaired hypoxia-induced mitochondrial clearance (mitophagy), which is believed to contribute to the pathogenesis of this disease[105]. Interestingly, mutations in other core splicing factors, including PRP31, which acts at the same step as PRP8, can also cause retinitis pigmentosa[106]. These examples illustrate how mutations affecting factors fundamental for the expression of most genes can nonetheless display very specific and rather defined phenotypes, once again arguing for strong effects of cellular context in the generation of disease phenotypes.

RBP-targeting therapeutics

In the past, RBPs were considered as largely ‘undruggable’, because of the lack of enzymatic pockets typically targeted by small molecules, the high structural similarity between individual members of RBD families, and the significant fraction of

unstructured regions present in these proteins. This picture is starting to change, and RBPs are emerging as promising novel therapeutic targets.

One example is provided by three families of natural products harbouring a common pharmacophore -spliceostatin, pladienolide B and GEX1- that display antitumour effects in various cancer models[107]. These compounds inhibit pre-mRNA splicing by binding to the interface between two interacting protein components of U2 small ribonucleoprotein complex (snRNP), SF3B1 and PHF5A[108,109] (Box 1). Although these interactions are essential for splicing of every intron, introns with different sequences flanking the branch site (and, consequently, harbouring different base-pairing potential with U2 small nuclear RNA) display differential sensitivity to these compounds. The resulting differential effects on splice site selection can form the basis of the anti-proliferative and pro-apoptotic effects of these drugs, for example, by modulating the balance between pro-apoptotic and anti-apoptotic isoforms of the protein MCL1 (refs 110-113). This example illustrates how drug targeting of an essential splicing step can yield effects on specific transcripts, with potential therapeutic utility. Interestingly, in addition to being the target of anti-tumour drugs, SF3B1 is frequently mutated in haematological malignancies as well as in solid tumours, leading to activation of cryptic 3' splice sites believed to alter the function of genes important for tumour progression (reviewed in refs 114 and 115).

A different class of RNA-focused therapy aims to block the access of RBPs to their binding sites on specific transcripts using chemically modified antisense oligonucleotides (ASOs)[116]. The recent success of this approach for the treatment of spinal muscular atrophy (SMA) paves the way to novel therapies for other genetic and multigenic pathologies[117]. SMA is a motor neuron disease caused by inactivation of the gene *SMN1*. Masking an intronic splicing silencer recognized by heterogeneous nuclear RNP (hnRNP) A1/A2 in pre-mRNAs of the closely related gene *SMN2* prevented binding of hnRNP A1/A2 to its cognate cis-acting regulatory sequence, and thus enhanced inclusion of the otherwise largely skipped exon 7, inducing the production of functional SMN protein and rescuing SMA-related phenotypes in mice[118,119]. This observation led to the development of nusinersen, a splicing-modifying ASO drug administered to patients with SMA as quarterly intrathecal lumbar injections, with remarkable therapeutic effects[120-123]. Of relevance, small molecules that promote inclusion of *SMN2* exon 7 have now also been identified[124-126], and some of these molecules are undergoing clinical trials as orally bioavailable drugs. Remarkably, these compounds act by promoting specific recognition by U1 snRNP of the 5' splice site associated with *SMN2* exon 7, demonstrating that it is possible to identify small-molecule modulators of a generic molecular recognition event with functional specificity and therapeutic relevance.

Another instance of 5' splice site regulation by small molecules with therapeutic relevance is familiar dysautonomia (OMIM #223900), a neurodegenerative genetic disorder that results from mutation of position +6 in intron 20 of the *IKBKAP* gene. The

mutation affects the interaction of the 5' splice site with U1 snRNP, leading to exon skipping and mRNA degradation by the nonsense mediated decay pathway, an effect which is particularly strong in the nervous system[127]. Several promising therapeutic compounds restore the use of the mutated 5' splice site through mechanisms that remain to be firmly established[127-130]. Whereas the plant cytokinin kinetin and the polyphenol epigallocatechin gallate (ECGC) seem to modulate the levels of particular splicing regulators, another compound (RECTAS) might inhibit the interaction of hnRNP H proteins with splicing silencers and, thus, facilitate binding of U1 snRNP to the mutated 5' splice site[129,131]. An approach conceptually similar to the mechanism of nusinersen, consisting of blocking the binding of hnRNP A1 to a splicing silencer, was also shown to restore splicing and IKBKAP function[132].

In sum, RNA–protein interactions can be regulated by small molecules that alter RNA-protein interactions or by ASOs that block the access of RBPs to their cognate binding sites, leading to specific effects that can be exploited for the treatment of disease. Other therapeutic approaches based on the use of oligonucleotides exist, including RNA aptamers (RNAs selected by in vitro evolution to specifically bind to target proteins) and short interfering RNAs, which trigger RNAi. Although only a fairly small number of oligonucleotide-based agents are currently in clinical use, advances in chemical modifications of their backbone or nucleobases, bioconjugation with lipids, sugars, peptides or antibodies, and a rapidly expanding catalogue of nanocarriers (including natural extracellular vesicles such as exosomes) hold the promise for overcoming current limitations in efficient tissue delivery (reviewed in ref. 116).

Conclusions and perspectives

This Review highlights the pervasive involvement of RBPs in human genetic disease. Who would have guessed that the number of disease-associated RBPs already exceeds 1,000? At the same time, it is apparent that, with very few exceptions, we lack pathomechanistic understanding. Pertinent questions to be answered include: what are the physiological roles of the ‘disease RBPs’, which complexes do they form and what are their dynamics? How do RBP mutations contribute to phenotypic outcomes at a mechanistic level? Which ‘target RNAs’ do RBPs recognize? Can we use structural data to aid in drug development? The vast genetic evidence calling for mechanistic, biochemical and structural insights highlights major unmet needs for intensive research in RNA biology, a widely overlooked opportunity for profound discoveries with translational potential and, consequently, a strong candidate for an urgent funding priority.

Even readers from the field of RNA biology may look sceptically at 3,000- 4,000 different human RBPs, especially when many of them lack the hallmarks of RBPs as we knew them a decade ago: architectural features such as RBDs and functional roles in the control of RNA fate. However, the fact that RNA aptamers can be selected as specific, high-affinity binders of proteins and other molecules[133] demonstrates that for a protein

to specifically interact with RNA, it does not necessarily need to harbour a well-established RBD.

Riboregulation[4,13], whereby RBPs are directly regulated by RNAs, offers an additional rationale for a much higher number of RBPs than previously anticipated: if one assumes that protein functions such as complex formation or enzymatic activity can be regulated by RNA, it follows that numerous proteins could bind RNA for regulatory purposes. Future work will have to uncover the scope of riboregulation in biology. As a corollary of this consideration, disease mutations in newly recognized RBPs need to be examined for their effects on RNA binding, especially when the phenotypic effect of the mutation cannot be plausibly explained on the basis of the previously known functions of the protein. An excellent example of this is hydroxysteroid dehydrogenase HSD17B10, a mitochondrial enzyme involved in the metabolism of sex hormones and neuroactive steroids. HSD17B10 is mutated in patients with a mitochondrial cardiomyopathy/neuropathy syndrome (OMIM #300438), and a non-enzymatic function has been suspected to explain the disease phenotype[134]. Following its recognition as an RBP, the major disease mutation in HSD17B10 was found to diminish the protein's binding to mitochondrial pre-tRNAs[9], explaining its role as a subunit of the RNase P complex that processes mitochondrial tRNAs[135]. Clearly, we need to approach the many RBPs involved in genetic disease with a widely open mind, ready to discover the unorthodox. Another recent example of the need for open-mindedness and for careful experimental analysis is the emerging importance of intrinsically unstructured regions in RNA binding, liquid de-mixing phenomena and their link to (neuro)degenerative disorders.

We also wish to draw attention to RNA biology and RBPs as a new therapeutic area, in spite of the complexities associated with proteins exerting broad biological functions and having numerous targets. The inherent potential for specific targeting by ASO is easy to appreciate, but the work required to show that ASOs reach their intracellular targets at concentrations required to elicit therapeutic benefits in real-day clinical situations is far from obvious. Likewise, one would not have expected effects on specific splice junctions and improved disease phenotypes from small-molecule drugs that target the core splicing apparatus. Therefore, we can only start to imagine the many opportunities that lie ahead.

Figure 1. RNA-binding proteins control RNA life. Nuclear (transcription, splicing, capping, polyadenylation) and cytoplasmic (transport, localization, translation, degradation) steps of mRNA metabolism are depicted. For a more detailed overview of the diversity of ribonucleoprotein particles (RNPs) present in the cell, see ref. 2. RNA Pol II, RNA polymerase II; RBP, RNA binding protein.

Figure 2. High-throughput approaches to identify RNA-binding proteins and their binding sites on RNA. Following the description of RNA interactome capture (RIC), a plethora of unbiased, high-throughput studies have ensued that identify RNA-binding proteins and their RNA-binding domains (RBDs). These methods include UV-crosslinking in living cells and, thereby, identify *bona fide in cellulo* interactions. **a)** Methods based on oligo(dT) capture. **RIC**: Proteins are covalently crosslinked to the RNA using UV irradiation, cells are disrupted, polyadenylated ((poly (A)) RNA is selected from cell extracts using oligo(dT) beads, contaminants are removed by rigorous washing, and associated proteins are identified by mass spectrometry (MS). This technology primarily identifies RBPs bound to mRNA and other poly(A) RNAs. Protocols using UV crosslinking at 254 nm (refs 6, 136), or feeding cells with the nucleoside analogue 4-thiouridine (4SU) followed by crosslinking at 365 nm (ref. 7) have been described. An improved version called enhanced RIC (**eRIC**) includes the use of LNA-modified oligo(dT) capture probes, among other changes, minimizing contamination by proteins bound to non-poly(A) RNAs[21]. Serial interactome capture (**serIC**) also allows higher specificity, as it includes two tandem RIC procedures separated by enzymatic digestion of contaminating DNA[25]. An extended RIC procedure termed **RBDmap** has been designed to identify RNA-binding sites on RBPs. This method consists of two oligo(dT) capture steps, and successive protease digestion of isolated ribonucleoprotein particles (RNPs) with two different proteases. In this technology, peptides *adjacent* to crosslinked peptides are identified[8,23]. A related method with a single oligo(dT) selection step termed **pCLAP** (peptide crosslinking and affinity purification) has been reported[24]. **b)** Methods based on click chemistry. Chemistry assisted RIC (**CARIC**): metabolic labelling of RNA using a 'clickable' uridine analogue (5-ethynyluridine (EU)) and 4SU is followed by 365 nm UV-crosslinking, addition of biotin to RNA using click chemistry, and the purification of RNP complexes with streptavidin beads. This method allows the purification of all newly synthesized RNA species and, thereby, is not limited to poly(A) RNA[26]. A similar method termed **RICK** (RNA interactome using click chemistry) lacking labelling with 4SU and, thereby, performing UV-crosslinking at 254 nm has also been reported[27]. **c)** Methods based on differential solubility. **OOPS** (orthogonal organic phase separation): this method is based on the principle that protein-RNA complexes, after covalent interactions promoted by UV-crosslinking, stay at the interphase of an aqueous-organic partition using acidic guanidinium thiocyanate-phenol-chloroform (also known as TRIzol). The method enables efficient recovery of all RNA species above 60 nucleotides and can be

used with fairly little input material[28]. Other methods termed **XRNAX**, **TRAPP** (total RNA-associated protein purification) and **2C** (complex capture) are also designed to exploit the chemical nature of crosslinked RNA-protein complexes for their selection irrespective of the polyadenylation status of the RNA[29,137,138]. **PTex** (phenol-toluol extraction) is based on similar principles using different reagents. First, RNP complexes accumulate in the aqueous phase after organic phenol-toluol separation; subsequent extraction with acidic phenol leaves RNP complexes in the interphase[30]. **d)** Methods to identify RBDs (see also RBDmap and pCLAP above). **RNPxl:** the first unbiased method to identify RNA binding sites in RBPs[31]. UV-crosslinked RNP complexes are digested with proteases and nucleases, and peptide-RNA conjugates are enriched by size exclusion chromatography (SEC) and reverse-phase C18 chromatography, followed by MS and a specialized computational workflow to identify the crosslinked amino acids and nucleotides. A related method called **RBR-ID** (RNA-binding region identification) uses RNA labelling with 4SU and crosslinking at 312 nm to detect reductions in MS intensity of crosslinked peptides compared with non-crosslinked controls[139]. **RBS-ID** (RNA binding site identification) uses hydrofluoride treatment to fully cleave the crosslinked RNA into mononucleotides, allowing direct MS identification of the crosslink at single amino acid resolution with high sensitivity[140]. **iTRAPP**, a modification of TRAPP where purified protein-RNA complexes are digested with nuclease P1 and trypsin followed by enrichment on titanium dioxide columns, also detects crosslinked peptides[137].

Figure 3. Human RNA-binding proteins involved in Mendelian and somatic genetic diseases. a) Detection frequency of RNA-binding proteins (RBPs). We expanded the collection of stringently curated human RNA interactome capture (RIC) studies[4] to 16 high-throughput studies in total[6-9,20-22,24-31,141], and the data are available in Tables S1, S2 and RBPbase. RBPs detected in at least two independent data sets are shown. Previously annotated RBPs show the highest frequency of detection. **b)** Human RBP superset. RBPs with at least 2 hits in RIC studies (2,650) were combined with a compilation of previously annotated RBPs from a curated list[2], Gene Ontology (GO)[142], RBPDB[143] and RNAcompete[144]. This results in a superset of 3,470 human RBPs. **c)** Human RBPs and RNA-binding domains (RBDs). Almost half of the proteins computationally predicted to contain an RBD display RNA-binding activity experimentally and are present in our human RBP superset. Disease-associated human RBPs **(**part **d)** and transcription factors (TFs) **(**part **e)** were annotated with Mendelian and somatic disease associations extracted from the Open Targets platform[145]. **f)** GO biological process enrichment analysis of disease-associated RBPs**. g, h)** Therapeutic areas of disease-associated RBPs. Disease mutations with an association score > 0.2 from Open Targets[145] were summed for selected therapeutic areas for RBPs, TFs[142,146], and all other proteins. RBPs that are also TFs were counted in both groups. The accumulated association scores were

normalized for the amount of proteins in each category. Additional information on methods is available at http://www.hentze.embl.de/public/hRBPdiseases.

Figure 4. Potential effects of mutations in RBP genes. Only mutations in the transcribed regions of genes have been considered. Mutations in non-coding regions (that is, untranslated regions (UTRs) and introns) or coding exons (open reading frames (ORFs)) of the pre-mRNA, may lead to altered mRNA levels, defective intracellular localization of the transcript or alternative transcript isoforms. For example, mutations in introns can lead to intron retention and subsequent nonsense-mediated decay of the transcript (*top*). Mutations in the coding region may elicit various effects depending on the type and location of the mutation. Nonsense mutations lead to protein truncation and defective activity. Missense mutations affecting sites of post-translational modification (PTM) may lead to altered signal perception. Mutations in RNA-binding domains (RBDs) or in protein-protein interaction (PPI) domains lead to defects in ribonucleoprotein particle (RNP) assembly and function. Mutations in low-complexity (LC) and disordered regions may lead to changes in the solubility of RNA-binding proteins (RBPs) ultimately resulting in accumulation of toxic aggregates. For those RBPs that display dual roles as enzymes, mutations in the enzymatic (Enz) domain may lead to defective catalysis (*bottom*). The consequences of mutations are not mutually exclusive. For instance, mutations in PTMs may not only lead to altered signal perception but also modify the interaction of the RBP with partners and its localization, stability or solubility.

Figure 5. Location and networks of disease-associated mutations. a) Location of disease-associated mutations in the mRNA of RNA-binding proteins (RBPs), transcription factors (TFs) and other proteins. Genomic locations of mutations were flattened and split into 5'untranslated region (UTR), 3'UTR, coding sequence (CDS) and intron regions. Genomic locations in PFAM domains were subtracted from these coordinates and divided into the RNA-binding domain (RBD) and other domains. TFs that bind RNA and contain RBDs have also been considered within this group. **b, c)** STRING[37] network of proteins with Mendelian or somatic disease-associated mutations for Alzheimer disease (part b) and for dyskeratosis congenita (part c). Red nodes indicate human RBPs. Size of nodes shows the disease association score; thickness of edges displays the evidence score from STRING. ABCA7, ATP-binding cassette subfamily A member 7; ACD, adrenocortical dysplasia protein homologue; ADAM10, disintegrin and metalloproteinase domain-containing protein 10; APOE, apolipoprotein E; APP, amyloid precursor protein; CSF1R, colony-stimulating factor 1 receptor; CTC1, conserved telomere maintenance component 1; DKC1, dyskerin; GRN, granulin precursor; HFE, homeostatic iron regulator; INPP4A, inositol polyphosphate-4-phosphatase type IA; MAPT, microtubule-associated protein Tau; MTND1, mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 1; NHP2, NHP2 ribonucleoprotein; NOP10, NOP10 ribonucleoprotein; PARN, poly(A)-specific ribonuclease; PSEN,

presenilin; RTEL1, regulator of telomere elongation helicase 1; SORL1, sortilin related receptor 1; TERT, telomerase reverse transcriptase; TINF2, TERF1 interacting nuclear factor 2; UNC5C, netrin receptor UNC5C; VCP, valosin-containing protein; WRAP53, WD repeat-containing protein antisense to TP53 gene.

Figure 6. RNA-binding proteins and phase transitions. a) Physical states of ribonucleoprotein particle (RNP) assemblies. Soluble RNPs can undergo initial de-mixing, resulting in liquid-like condensates that may further transit into more viscous hydrogel-like states and solid-like pathological aggregates. Dynamic RNP assemblies within the cell are thought to interchange between the first two states, whereas the last state is largely irreversible and toxic. **b)** Principles of FUS (fused in sarcoma) liquid de-mixing. Domain organization of FUS (*top*). Hotspots for mutations causing the neurodegenerative disorders amyotrophic lateral sclerosis and frontotemporal dementia are indicated. Contacts between tyrosine residues in the prion-like domain (PrLD) and arginine residues in the region rich in arginine and glycine residues (RGG) promote liquid–liquid phase separation, with mutations promoting aberrant transitions to toxic aggregates (*bottom*). NES, nuclear export signal; NLS, nuclear localization signal; RRM, RNA recognition motif; ZF, zinc finger.

Figure 7. Mechanisms of disease in fragile X syndrome and FXTAS. a) Domain organization of FMRP, indicating the location of the disease-causing mutation I304N. **b)** Trinucleotide expansions ($[CGG]_n$) in the FMR1 gene are viewed as a thin (fragile) region of the X-chromosome. Expansions to 55-200 repeats lead to the formation of nuclear aggregates containing polyG-FMRP or excess FMR1 mRNA that sequesters other RNA-binding proteins, leading to fragile X-associated tremor ataxia syndrome (FXTAS). Expansions over 230 repeats lead to hypermethylation and silencing of the FMR1 gene, causing fragile X syndrome. FMRP has been proposed to regulate mRNA translation in neurons by various mechanisms. One of these, repression of translation elongation, has gathered more attention and is depicted here. FMRP may also aid in the transport of ribonucleoprotein particle granules towards synaptic terminals, where mRNAs are derepressed in a stimulus-dependent manner. KH, K-homology; NES, nuclear export signal; NLS, nuclear localization signal; RGG, region rich in arginine and glycine residues.

Box 1. The splicing process and current pharmacological targets

Removal of introns from mRNA precursors (pre-mRNA splicing) requires recognition of sequences at the 5' and 3' ends of the intron (splice sites) by U1 and U2 small nuclear ribonucleoprotein complexes (snRNPs), respectively (see the figure). Recognition involves base-pairing interactions between the pre-mRNA and RNA components of the snRNPs: U1 small nuclear RNA (snRNA) base pairs with the 5' splice site and U2 snRNAs with the branch site region closely upstream of the 3' splice site. The branch site is an

adenosine residue that forms a 2'-5' phosphodiester bond with the 5' splice site during the first catalytic step of the splicing process. As splice site sequences of higher eukaryotes are relatively variable, the configuration of base-pairing interactions (number and position of base-paired nucleotides) varies among splice sites. Small molecules that stabilize recognition of specific 5' splice sites by U1 snRNP are being tested as orally bioavailable drugs for the treatment of spinal muscular atrophy (SMA). Antisense oligonucleotides (ASOs) that inhibit the binding of heterogeneous nuclear ribonucleoproteins (hnRNPs) to an intronic splicing silencer have also been shown effective against SMA (see main text for details). Small molecules that bind to the interface between the U2 snRNP components SF3B1 and PHF5A display antitumour properties. These molecules prevent a conformational change in the HEAT repeats domain of SF3B1 that is essential for proper recognition of the branch site–U2 snRNA interaction[108,109,147].

Glossary:

RNA interactome- The group of proteins that interact with RNA.

Intrinsically disordered regions- Protein regions that lack a stable secondary or tertiary structure.

Germline mutation- A mutation that occurs in the germline and, therefore, is inherited by the next generation.

Somatic mutations- Mutations that occur in somatic cells and, therefore, are not transmitted to the next generation.

Genetic disease- A disease that is caused by mutations in a gene or a group of genes.

Mendelian disorders- Diseases caused by gene mutations in the germline and inherited either in a dominant or recessive manner according to Mendelian laws.

References

1. Mitchell, S.F. & Parker, R. Principles and properties of eukaryotic mRNPs. *Mol Cell.* **54**, 547-558 (2014).

2. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829-845 (2014).

3. Singh, G., Pratt, G., Yeo, G.W. & Moore, M.J. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. *Annu. Rev. Biochem*. **84**, 325-354 (2015).

4. Hentze, M.W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell. Biol*. **19**, 327-341 (2018).

5. Bou-Nader, C., Gordon, J.M., Henderson, F.E. and Zhang, J. The search for a PKR code-differential regulation of protein kinase R activity by diverse RNA and protein regulators. *RNA* **25**, 539-556 (2019).

6. Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406 (2012).
This is one of the two pioneer studies reporting the unbiased identification of RNA binding proteins, performed in Hela cells.

7. Baltz, A.G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674-690 (2012).
This is one of the two pioneer studies reporting the unbiased identification of RNA binding proteins, performed in HEK293 cells.

8. Castello, A. et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696-710 (2016).
This article describes RBDmap, a method for the high-throughput identification of RBDs within RBPs. Peptides adjacent to the actual RNA binding residues are identified.

9. Beckmann, B.M. et al. The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun*. **6**, 10127 (2015).

10. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434-448 (2008).

11. Iadevaia, V. & Gerber, A. P. Combinatorial Control of mRNA Fates by RNA-Binding Proteins and Non-Coding RNAs. *Biomolecules* **5**, 2207-2222 (2015).

12. Achsel, T. & Bagni, C. Cooperativity in RNA-protein interactions: the complex is more than the sum of its partners. *Curr. Opin. Neurobiol*. **39**, 146-151 (2016).

13. Beckmann, B. M., Castello, A. & Medenbach, J. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Arch* **468**, 1029-1040 (2016).

14. Morris, A. R., Mukherjee, N. & Keene, J. D. Systematic analysis of posttranscriptional gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 162-180 (2010).

15. Imig, J., Kanitz, A. & Gerber, A. P. RNA regulons and the RNA-protein interaction network. *Biomol. Concepts* **3**, 403-414 (2012).

16. Castello, A., Fischer, B., Hentze, M.W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet*. **29**, 318-327 (2013).

17. Gerstberger, S., Hafner, M., Ascano, M. & Tuschl, T. Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. *Adv. Exp. Med. Biol.* **825**, 1-55 (2014).

18. De Conti, L., Baralle, M. & Buratti, E. Neurodegeneration and RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* **8**, e1394 (2017).

19. Conlon, E. G. & Manley, J. L. RNA-binding proteins in neurodegeneration: mechanisms in aggregate. *Genes Dev*. **31**, 1509-1528 (2017).

20. Garcia-Moreno, M. et al. System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Mol Cell* **74**, 196–211 (2019).

21. Perez-Perri, J. I. et al. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun*. **9**, 4408 (2018).

22. Backlund, M. et al. Plasticity of nuclear and cytoplasmic stress responses of RNA-binding proteins. *Nucleic Acids Res*. **48**, 4725–4740 (2020).

23. Castello, A. et al. Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. *Nat. Protoc*. **12**, 2447-2464 (2017).

24. Mullari, M., Lyon, D., Jensen, L. J. & Nielsen, M. L. Specifying RNA-Binding Regions in Proteins by Peptide Cross-Linking and Affinity Purification. *J. Proteome Res*. **16**, 2762-2772 (2017).

25. Conrad, T. et al. Serial interactome capture of the human cell nucleus. *Nat Commun*. **7**, 11212 (2016).

26. Huang, R., Han, M., Meng, L. & Chen, X. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proc. Natl. Acad. Sci. USA* **115**, E3879-E3887 (2018).

27. Bao, X. et al. Capturing the interactome of newly transcribed RNA. *Nat. Methods* **15**, 213-220 (2018).

28. Queiroz, R. M. L. et al. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat. Biotechnol*. **37**, 169-178 (2019).

29. Trendel, J. et al. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403 (2019).

30. Urdaneta, E. C. et al. Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat. Commun.* **10**, 990 (2019).

31. Kramer, K. et al. Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11**, 1064-1070 (2014).
This article describes the first unbiased method to identify RNA binding sites in RBPs.

32. Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci.* **16**, 323-326 (1991).

33. Wang, Z. L. et al. Comprehensive Genomic Characterization of RNA-Binding Proteins across Human Cancers. *Cell Rep*. **22**, 286-298 (2018).

34. Wang, X. et al. UDP-glucose accelerates SNAI1 mRNA decay and impairs lung cancer metastasis. *Nature* **571**, 127-131 (2019).
This article shows the exquisite regulation of the RBP HuR by metabolites, and the consequences for selective expression of its target Snail and tumorigenesis.

35. Pavitt, G. D. Regulation of translation initiation factor eIF2B at the hub of the integrated stress response. *Wiley Interdiscip. Rev. RNA* **9**, e1491 (2018).

36. Beaumont, C. et al. Mutation in the iron responsive element of the L ferritin mRNA in a family with dominant hyperferritinaemia and cataract. *Nat. Genet*. **11**, 444-446 (1995).

37. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. **47**, D607–D613 (2019).

38. Kato, M. et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).
This paper is the first report (together with an accompanying paper from the same lab) showing that low complexity domains within RBPs can form liquid-like gels.

39. Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208-219 (2015).

40. Aguzzi, A. & Altmeyer, M. Phase Separation: Linking Cellular Compartmentalization to Disease. *Trends Cell Biol.* **26**, 547-558 (2016).

41. Alberti, S. Phase separation in biology. *Curr. Biol.* **27**, R1097-R1102 (2017).

42. Guillén-Boixet, J. et al. RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361.e17 (2020).
Together with Yang et al. and Sanders et al., this article shows the principles of stress granule formation by liquid-liquid phase separation.

43. Yang, P. et al. G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345.e28 (2020).

44. Sanders, D. W. et al. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324.e28 (2020).

45. Mittag, T. & Parker, R. Multiple Modes of Protein-Protein Interactions Promote RNP Granule Assembly. *J. Mol. Biol*. **430**, 4636-4649 (2018).

46. McSwiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev*. **33**, 1619-1634 (2019).

47. Patel, A. et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-1077 (2015).
This article provides a rationale for disorders characterized by accumulation of amyloid fibers

48. Bogaert, E. et al. Molecular Dissection of FUS Points at Synergistic Effect of Low-Complexity Domains in Toxicity. *Cell Rep.* **24**, 529-537 (2018).

49. Wang, J. et al. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699 (2018).
This work describes the properties of a series of amino acids (Y, R, G, Q, S) in promoting or counteracting phase separation, and provides a model to predict the phase-separation tendencies of RBPs based on their sequence.

50. Dormann, D. et al. ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J*. **29**, 2841-2857 (2010).

51. Naumann, M. et al. Impaired DNA damage response signaling by FUS-NLS mutations leads to neurodegeneration and FUS aggregate formation. *Nat Commun*. **9**, 335 (2018).

52. Alberti, S. & Hyman, A. A. Are aberrant phase transitions a driver of cellular aging? *Bioessays* **38**, 959-968 (2016).

53. Verkerk, A. J. et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905-914 (1991).

54. Chen, E. & Joseph, S. Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie* **114**, 147-154 (2015).

55. Dictenberg, J. B., Swanger, S. A., Antar, L. N., Singer, R. H. & Bassell, G. J. A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev. Cell* **14**, 926-939 (2008).

56. Napoli, I. et al. The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* **134**, 1042-1054 (2008).

57. Darnell, J. C. et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-261 (2011).

58. De Boulle, K. et al. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet*. **3**, 31-35 (1993).

59. Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L. & Dreyfuss, G. Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* **77**, 33-39 (1994).

60. Feng, Y. et al. FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol. Cell* **1**, 109-118 (1997).

61. Darnell, J. C. et al. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489-499 (2001).

62. Darnell, J. C. et al. Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev*. **19**, 903-918 (2005).

63. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K. & Joseph, S. Fragile X mental retardation protein regulates translation by binding directly to the ribosome. *Mol. Cell* **5**4, 407-417 (2014).
This report shows the cryo-EM reconstruction of FMRP bound to the 80S ribosome and provides a rationale for inhibition of translation elongation by this protein.

64. Ascano, M. Jr. et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382-386 (2012).

65. Bechara, E. G. et al. A novel function for fragile X mental retardation protein in translational activation. *PLoS Biol*. **7**, e16 (2009).

66. Zalfa, F. et al. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* **112**, 317-327 (2003).

67. Jin, P. et al. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci.* **7**, 113-117 (2004).

68. Edbauer, D. et al. Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* **65**, 373-384 (2010).

69. Muddashetty, R. S. et al. Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. *Mol. Cell* **42**, 673-688 (2011).

70. Greenblatt, E. J. & Spradling, A. C. Fragile X mental retardation 1 gene enhances the translation of large autism-related proteins. *Science* **361**, 709-712 (2018).

71. Tassone, F., Iwahashi, C. & Hagerman, P. J. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol*. **1**, 103-105 (2004).

72. Greco, C. M. et al. Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain* **129**, 243–255 (2006).

73. Jin, P. et al. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* **39**, 739-747 (2003).

74. Jin, P. et al. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron* **55**, 556-564 (2007).

75. Sofola, O. A. et al. RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* **55**, 565-571 (2007).

76. Sellier, C. et al. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J*. **29**, 1248-1261 (2010).

77. Todd, P. K. et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455 (2013).

78. Sellier, C. et al. Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* **93**, 331-347 (2017).

79. Johnston, J. J. et al. Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate. *Am. J. Hum. Genet.* **86**, 743-748 (2010).

80. Gripp, K. W. et al. Long-term survival in TARP syndrome and confirmation of RBM10 as the disease-causing gene. *Am. J. Med .Genet. A*. **155A**, 2516-2520 (2011).

81. Wang, Y. et al. Integrative analysis revealed the molecular mechanism underlying RBM10-mediated splicing regulation. *EMBO Mol. Med.* **5**, 1431-1442 (2013).

82. Rodor, J., FitzPatrick, D. R., Eyras, E. & Cáceres, J. F. The RNA-binding landscape of RBM10 and its role in alternative splicing regulation in models of mouse early development. *RNA Biol*. **14**, 45-57 (2017).

83. Mohan, N., Kumar, V., Kandala, D. T., Kartha, C. C. & Laishram, R. S. A Splicing-Independent Function of RBM10 Controls Specific 3' UTR Processing to Regulate Cardiac Hypertrophy. *Cell Rep*. **24**, 3539-3553 (2018).

84. Imielinski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107-1120 (2012).
This article uncovers genetic evidence that RNA binding proteins involved in splicing, including RBM10, play a role in lung cancer progression.

85. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-50 (2014).

86. Yin, L. L. et al. A gene mutation in RNA-binding protein 10 is associated with lung adenocarcinoma progression and poor prognosis. *Oncol. Lett.* **16**, 6283-6292 (2018).

87. Chen, H. et al. Genomic and immune profiling of pre-invasive lung adenocarcinoma. *Nat. Commun.* **10**, 5472 (2019).

88. Giannakis, M. et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep.* **15**, 857-865 (2016).

89. Witkiewicz, A. K. et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun*. **6**, 6744 (2015).

90. Seiler, M. et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep*. **23**, 282-296 (2018).

91. Bechara, E. G., Sebestyén, E., Bernardis, I., Eyras, E. & Valcárcel, J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell*. **52**, 720-733 (2013).

92. Jin, X. et al. RBM10 inhibits cell proliferation of lung adenocarcinoma via RAP1/AKT/CREB signalling pathway. *J. Cell. Mol. Med*. **23**, 3897-3904 (2019).

93. Loiselle, J. J., Roy, J. G. & Sutherland, L. C. RBM10 promotes transformation-associated processes in small cell lung cancer and is directly regulated by RBM5. *PLoS One* **12**, e0180258 (2017).

94. Loiselle, J. J. & Sutherland, L. C. J. RBM10: Harmful or helpful-many factors to consider. *Cell Biochem.* **119**, 3809-3818 (2018).

95. Hernández, J. et al. Tumor suppressor properties of the splicing regulatory factor RBM10. *RNA Biol*. **13**, 466-472 (2016).

96. Zhao, J. et al. Functional analysis reveals that RBM10 mutations contribute to lung adenocarcinoma pathogenesis by deregulating splicing. *Sci. Rep*. **7**, 40488 (2017).

97. Collins, K. M. et al. An RRM-ZnF RNA recognition module targets RBM10 to exonic sequences to promote exon exclusion. *Nucleic Acids Res*. **45**, 6761-6774 (2017).

98. Grainger, R. J. & Beggs, J. D. Prp8 protein: at the heart of the spliceosome. *RNA* **11**, 533-557 (2005).

99. Maeder, C., Kutach, A. K. & Guthrie, C. ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. *Nat. Struct. Mol. Biol*. **16**, 42-48 (2009).

100. Mozaffari-Jovin, S. et al. Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341**, 80-84 (2013).

101. Yan, C. et al. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182-1191 (2015).

102. McKie, A. B. et al. Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.* **10**, 1555-1562 (2001).

103. Pena, V., Liu, S., Bujnicki, J. M., Lührmann, R. & Wahl, M. C. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol. Cell* **25**, 615-624 (2007).
This paper shows that amino acid mutations in a domain of the core splicing protein PRP8 which disrupt interactions with the RNA helicase BRR2 involved in splicing catalysis, underlie the eye degeneration disease retinitis pigmentosa.

104. Farkas, M. H. et al. Mutations in pre-mRNA processing factors 3, 8, and 31 cause dysfunction of the retinal pigment epithelium. *Am. J. Pathol*. **184**, 2641-2652 (2014).

105. Xu, G. et al. Autosomal dominant retinitis pigmentosa-associated gene PRPF8 is essential for hypoxia-induced mitophagy through regulating ULK1 mRNA splicing. *Autophagy* **14**, 1818-1830 (2018).

106. Buskin, A. et al. Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. *Nat. Commun*. **9**, 4234 (2018).

107. Desterro, J., Bak-Gordon, P. & Carmo-Fonseca, M. Targeting mRNA processing as an anticancer strategy. *Nat. Rev. Drug Discov*. **19**, 112–129 (2020).

108. Cretu, C. et al. Structural basis of splicing modulation by antitumor macrolide compounds. *Mol. Cell* **70**, 265-273 (2018).
This article shows that certain anti-tumor drugs inhibit a conformational change required for recognition of the pre-mRNA branch site by the splicing factor SF3B1.

109. Finci, L. I. et al. The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. *Genes Dev.* **32**, 309-320 (2018).

110. Corrionero, A., Miñana, B. & Valcárcel, J. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. *Genes Dev*. **25**, 445-459 (2011).

111. Folco, E. G., Coil, K. E. & Reed, R. The anti-tumor drug E7107 reveals an essential role for SF3b in remodeling U2 snRNP to expose the branch point-binding region. *Genes Dev.* **25**, 440-444 (2011).

112. Gao, Y. & Koide, K. Chemical perturbation of Mcl-1 pre-mRNA splicing to induce apoptosis in cancer cells. *ACS Chem. Biol.* **8**, 895-900 (2013).

113. Vigevani, L., Gohr, A., Webb, T., Irimia, M. & Valcárcel, J. Molecular basis of differential 3' splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun*. **8**, 2100 (2017).

114. Rahman, M. A., Krainer, A. R. & Abdel-Wahab, O. SnapShot: Splicing Alterations in Cancer. *Cell* **180**, 208-208.e1 (2020).

115. Bonnal, S. C., López-Oreja, I. & Valcárcel, J. Roles and mechanisms of alternative splicing in cancer - implications for care. *Nat Rev Clin Oncol.* **17**, 457-474 (2020).

116. Roberts, T. C., Langer, R. & Wood, M. J. A. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov.* Aug 11, 1-22 (2020).

117. Bennett, C. F. Therapeutic antisense oligonucleotides are coming of age. *Annu. Rev. Med.* **70**, 307–321 (2019).

118. Hua, Y., Vickers. T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet*. **82**, 834-848 (2008).

119. Hua, Y. et al. Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev*. **24**, 1634-1644 (2010).
This paper provides the first evidence that splicing-modifying ASOs can have therapeutic effects in a mouse model of SMA.

120. Finkel, R. S. et al. Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N. Engl. J. Med*. **377**, 1723-1732 (2017).
This paper presents the first successful clinical trial of a splicing-modifying ASO for the treatment of patients with SMA.

121. Mercuri, E. et al. Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy. *N. Engl. J. Med*. **378**, 625-635 (2018).

122. Pane, M. et al. Nusinersen in type 1 SMA infants, children and young adults: Preliminary results on motor function. *Neuromuscul. Disord*. **28**, 582-585 (2018).

123. De Vivo, D. C. et al. Nusinersen initiated in infants during the presymptomatic stage of spinal muscular atrophy: Interim efficacy and safety results from the Phase 2 NURTURE study. *Neuromuscul. Disord.* **29**, 842-856 (2019).

124. Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol*. **11**, 511-517 (2015).

125. Sivaramakrishnan, M. et al. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun*. **8**, 1476 (2017).
This article provides structural insights into how a small molecule can modulate the recognition of a specific 5' splice site by U1 snRNP with potential therapeutic effects.

126. García-López, A. et al. Targeting RNA structure in SMN2 reverses spinal muscular atrophy molecular phenotypes. *Nat. Commun*. **9**, 2032 (2018).

127. Slaugenhaupt, S. A. et al. Tissue-specific expression of a splicing mutation in the IKBKAP gene causes familial dysautonomia. *Am. J. Hum. Genet.* **68**, 598-605 (2001).

128. Anderson, S. L., Qiu, J. & Rubin, B. Y. EGCG corrects aberrant splicing of IKAP mRNA in cells from patients with familial dysautonomia. *Biochem. Biophys. Res. Commun*. **310**, 627-633 (2003).

129. Yoshida, M. et al. Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. *Proc. Natl. Acad. Sci. USA* **112**, 2764-2769 (2015).

130. Morini, E. et al. ELP1 Splicing Correction Reverses Proprioceptive Sensory Loss in Familial Dysautonomia. *Am. J. Hum. Genet*. **104**, 638-650 (2019).

131. Swanson, M. S. Rectifying RNA splicing errors in hereditary neurodegenerative disease. *Proc. Natl. Acad. Sci. USA* **112**, 2637-2638 (2015).

132. Bruun, G. H. et al. Blocking of an intronic splicing silencer completely rescues IKBKAP exon 20 splicing in familial dysautonomia patient cells. *Nucleic Acids Res.* **46**, 7938-7952 (2018).

133. Kaur, H., Bruno, J. G., Kumar, A. & Sharma, T.K. Aptamers in the therapeutics and diagnostics pipelines. *Theranostics* **8**, 4016-4032 (2018).

134. Rauschenberger, K. et al. A non-enzymatic function of 17beta-hydroxysteroid dehydrogenase type 10 is required for mitochondrial integrity and cell survival. *EMBO Mol. Med.* **2**, 51–62 (2010)].

135. Holzmann, J. et al. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).

136. Castello, A. et al. System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.* **8**, 491-500 (2013).

137. Shchepachev, V. et al. Defining the RNA interactome by total RNA-associated protein purification. *Mol. Syst. Biol*. **15**, e8689 (2019).

138. Asencio, C., Chatterjee, A. & Hentze, M.W. Silica-based solid-phase extraction of cross-linked nucleic acid-bound proteins. *Life Sci Alliance* **1**, e201800088 (2018).

139. He, C. et al. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol. Cell* **64**, 416-430 (2016).
This article describes RBR-ID, a method for the high-throughput identification of RNA binding peptides within RBPs. The peptide that actually crosslinks to the RNA is identified.

140. Bae, J.W., Kwon, S.C., Na, Y., Kim, V.N. & Kim, J. Chemical RNA digestion enables robust RNA-binding site mapping at single amino acid resolution. *Nat. Struc. Mol. Biol.* (2020) doi: 10.1038/s41594-020-0436-2.
This article describes a method for the sensitive, high-throughput identification of RNA binding amino acids within RBPs.

141. Treiber, T. et al. A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. *Mol Cell* **66**, 270–284.e13 (2017).

142. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. **47**, D330–D338 (2019).

143. Cook, K. B., Kazan, H., Zuberi, K., Morris, Q. & Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res*. **39**, D301–8 (2011).

144. Ray, D. et al. Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol.* **27**, 667–670 (2009).

145. Carvalho-Silva, D. et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res*. **47**, D1056–D1065 (2019).

146. Lambert, S. A. et al. The Human Transcription Factors. *Cell* **175**, 598–599 (2018).

147. Teng, T. et al. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. *Nat. Commun.* **8**, 15522 (2017).

Acknowledgements
The authors thank J. Guillén-Boixet for insightful comments and suggestions on the manuscript and S. Sahadevan for help with parsing Open Targets data. The authors acknowledge funding from the Spanish Ministry of Science and Innovation (MICINN, PGC2018-099697-B-I00 to F.G. and BFU2017-89308-P to J.V.), 'la Caixa' Foundation (ID 100010434 under the agreement LCF/PR/HR17/52150016 to F.G.), the Catalan

Government (2017SGR534 to F.G. and J.V.), the ERC (AdvG 670146 to J.V.), as well as the EMBL Partnership, the Severo Ochoa and CERCA Programs.

Author contributions

T.S. performed the bioinformatics analyses. All authors contributed to all other aspects of the article.

Data availability

The data that support the findings of this study are available as Supplementary Tables and in RBPbase: https://rbpbase.shiny.embl.de/.

Code availability

The source code used in this paper and written by T.S. ispublicly available at http://www.hentze.embl.de/public/hRBPdiseases/.

Related links

OMiM #223900: https://omim.org/entry/223900
OMiM #300438: https://omim.org/entry/300438
OMiM #311900: https://omim.org/entry/311900
OMiM #600886: https://omim.org/entry/600886
OMiM #603896: https://omim.org/entry/603896
Open Targets: https://www.opentargets.org/
RBPbase: http://rbpbase.shiny.embl.de

Figure 1

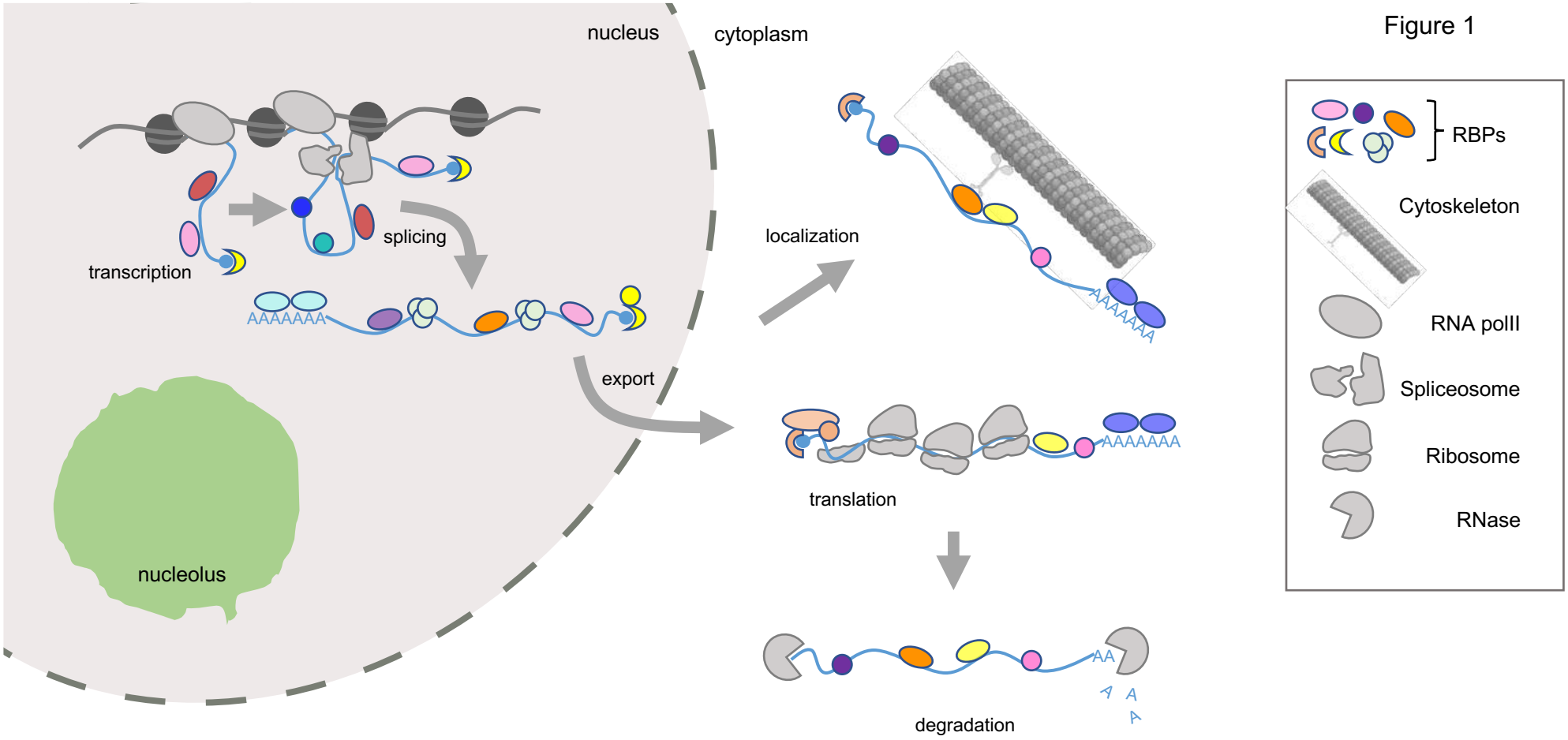

nucleus
cytoplasm
transcription
splicing
AAAAAAA
export
localization
AAAAAAA
translation
AAAAAAA
nucleolus
AA
degradation
RBPs
Cytoskeleton
RNA polII
Spliceosome
Ribosome
RNase

Figure 2

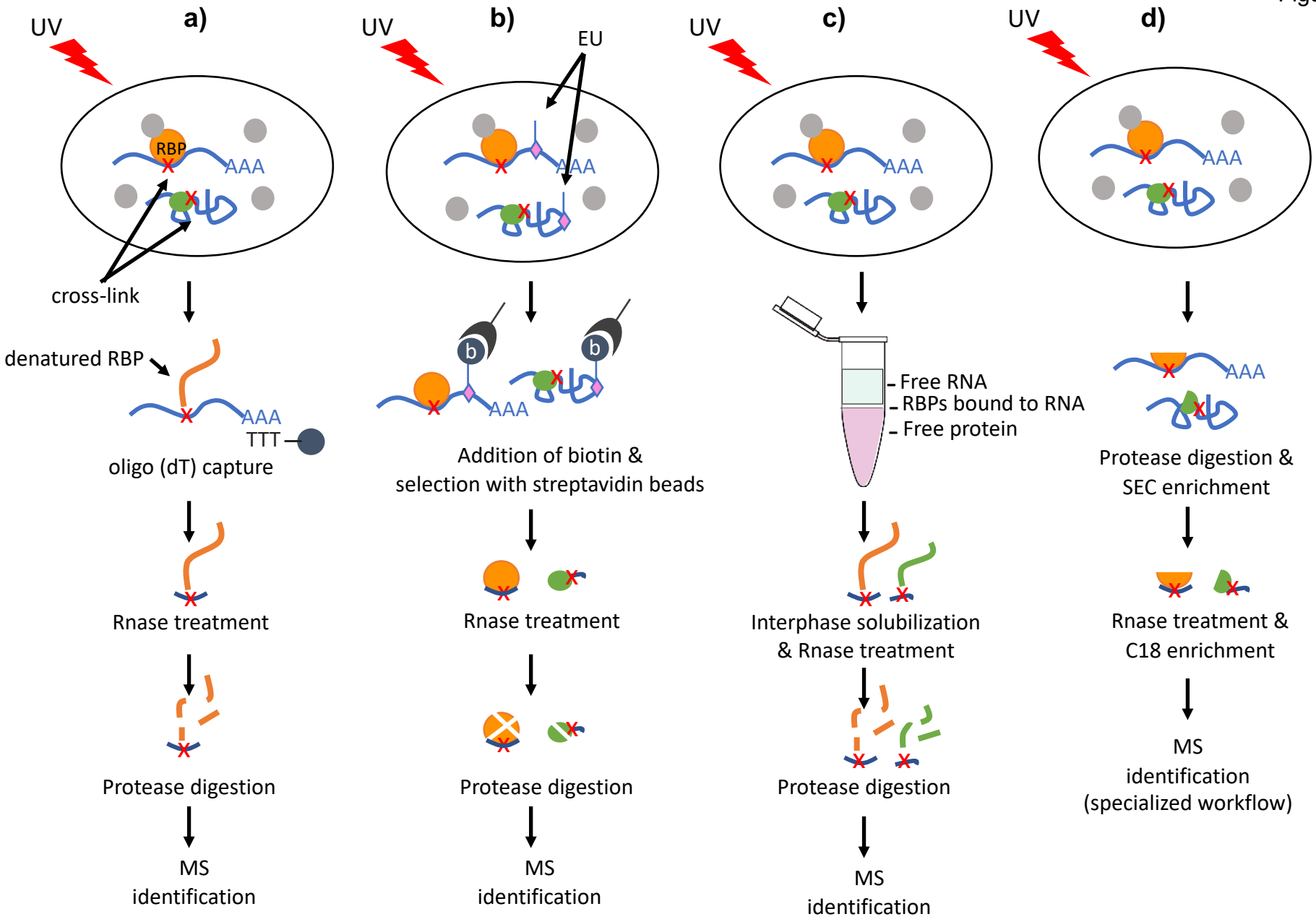

a)
UV
RBP
AAA
cross-link
denatured RBP
AAA
TTT
oligo (dT) capture
Rnase treatment
Protease digestion
MS
identification
b)
UV
EU
AAA
b
b
AAA
Addition of biotin &
selection with streptavidin beads
Rnase treatment
Protease digestion
MS
identification
c)
UV
AAA
Free RNA
RBPs bound to RNA
Free protein
Interphase solubilization
& Rnase treatment
Protease digestion
MS
identification
d)
UV
AAA
AAA
Protease digestion &
SEC enrichment
Rnase treatment &
C18 enrichment
MS
identification
(specialized workflow)

Figure 3

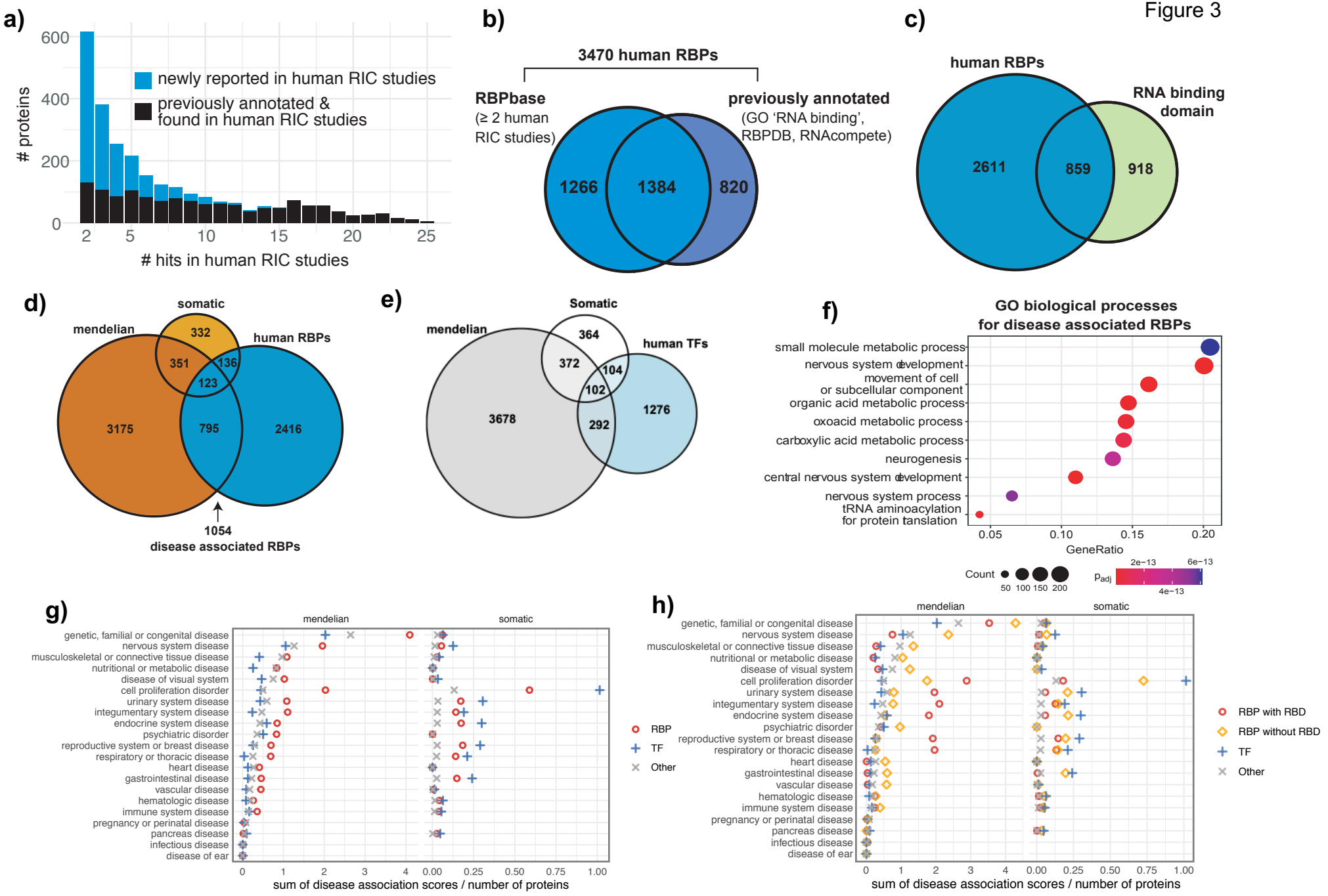

a)
proteins
600
400
200
0
newly reported in human RIC studies
previously annotated & found in human RIC studies
2
5
10
15
20
25
hits in human RIC studies
b)
3470 human RBPs
RBPbase
(≥ 2 human RIC studies)
previously annotated
(GO 'RNA binding', RBPDB, RNAcompete)
1266
1384
820
c)
human RBPs
RNA binding domain
2611
859
918
d)
somatic
mendelian
human RBPs
332
351
136
123
3175
795
2416
1054
disease associated RBPs
e)
Somatic
mendelian
human TFs
364
372
104
102
3678
292
1276
f)
GO biological processes for disease associated RBPs
small molecule metabolic process
nervous system development
movement of cell or subcellular component
organic acid metabolic process
oxoacid metabolic process
carboxylic acid metabolic process
neurogenesis
central nervous system development
nervous system process
tRNA aminoacylation for protein translation
0.05
0.10
0.15
0.20
GeneRatio
Count
50
100
150
200
p_{adj}
2e−13
6e−13
4e−13
g)
mendelian
somatic
genetic, familial or congenital disease
nervous system disease
musculoskeletal or connective tissue disease
nutritional or metabolic disease
disease of visual system
cell proliferation disorder
urinary system disease
integumentary system disease
endocrine system disease
psychiatric disorder
reproductive system or breast disease
respiratory or thoracic disease
heart disease
gastrointestinal disease
vascular disease
hematologic disease
immune system disease
pregnancy or perinatal disease
pancreas disease
infectious disease
disease of ear
0
1
2
3
4
0.00
0.25
0.50
0.75
1.00
sum of disease association scores / number of proteins
RBP
TF
Other
h)
mendelian
somatic
genetic, familial or congenital disease
nervous system disease
musculoskeletal or connective tissue disease
nutritional or metabolic disease
disease of visual system
cell proliferation disorder
urinary system disease
integumentary system disease
endocrine system disease
psychiatric disorder
reproductive system or breast disease
respiratory or thoracic disease
heart disease
gastrointestinal disease
vascular disease
hematologic disease
immune system disease
pregnancy or perinatal disease
pancreas disease
infectious disease
disease of ear
0
1
2
3
4
0.00
0.25
0.50
0.75
1.00
sum of disease association scores / number of proteins
RBP with RBD
RBP without RBD
TF
Other

Figure 4

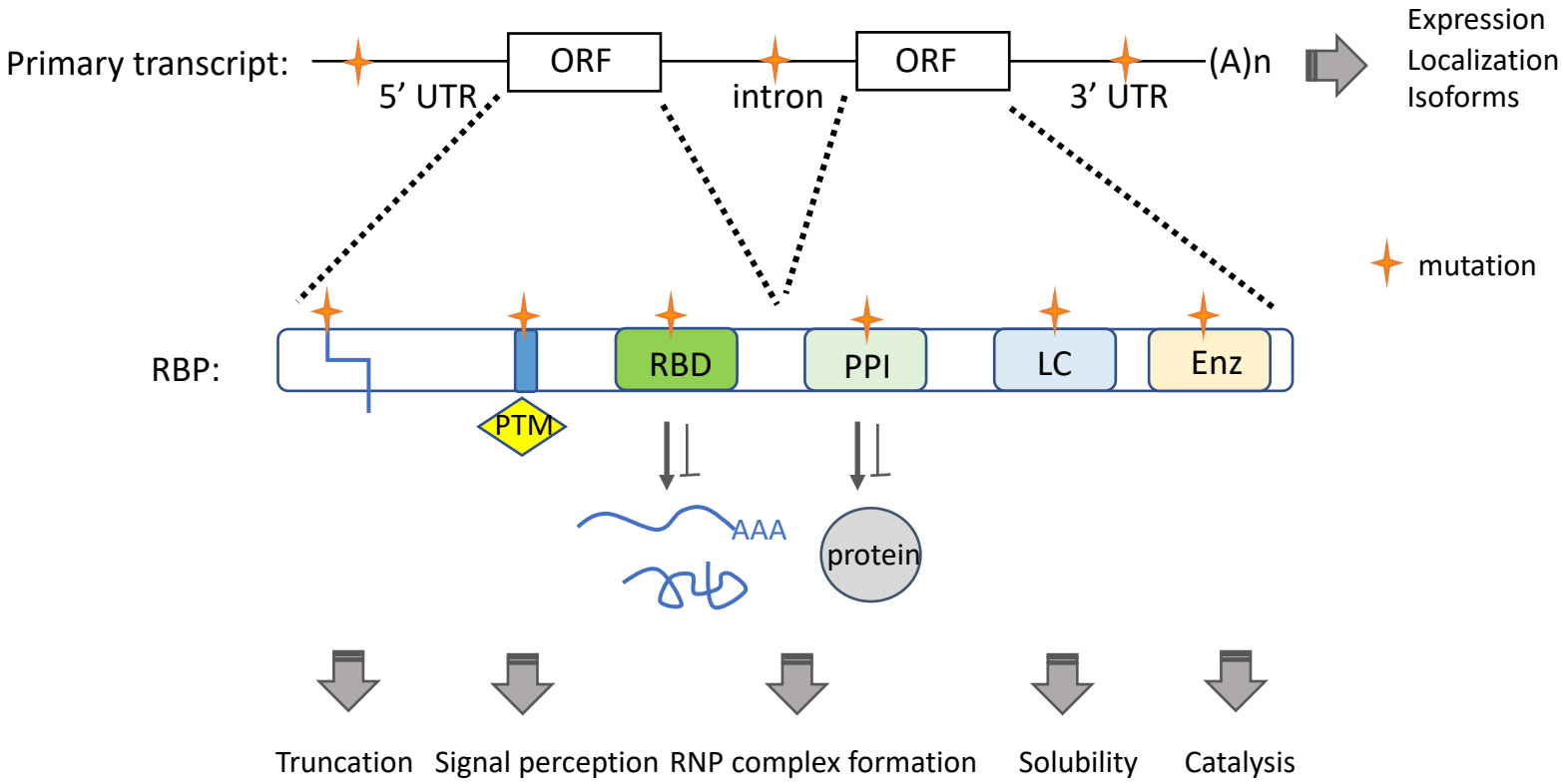

Primary transcript:
5' UTR
ORF
intron
ORF
3' UTR
(A)n
Expression
Localization
Isoforms
mutation
RBP:
RBD
PPI
LC
Enz
PTM
AAA
protein
Truncation
Signal perception
RNP complex formation
Solubility
Catalysis

Figure 5

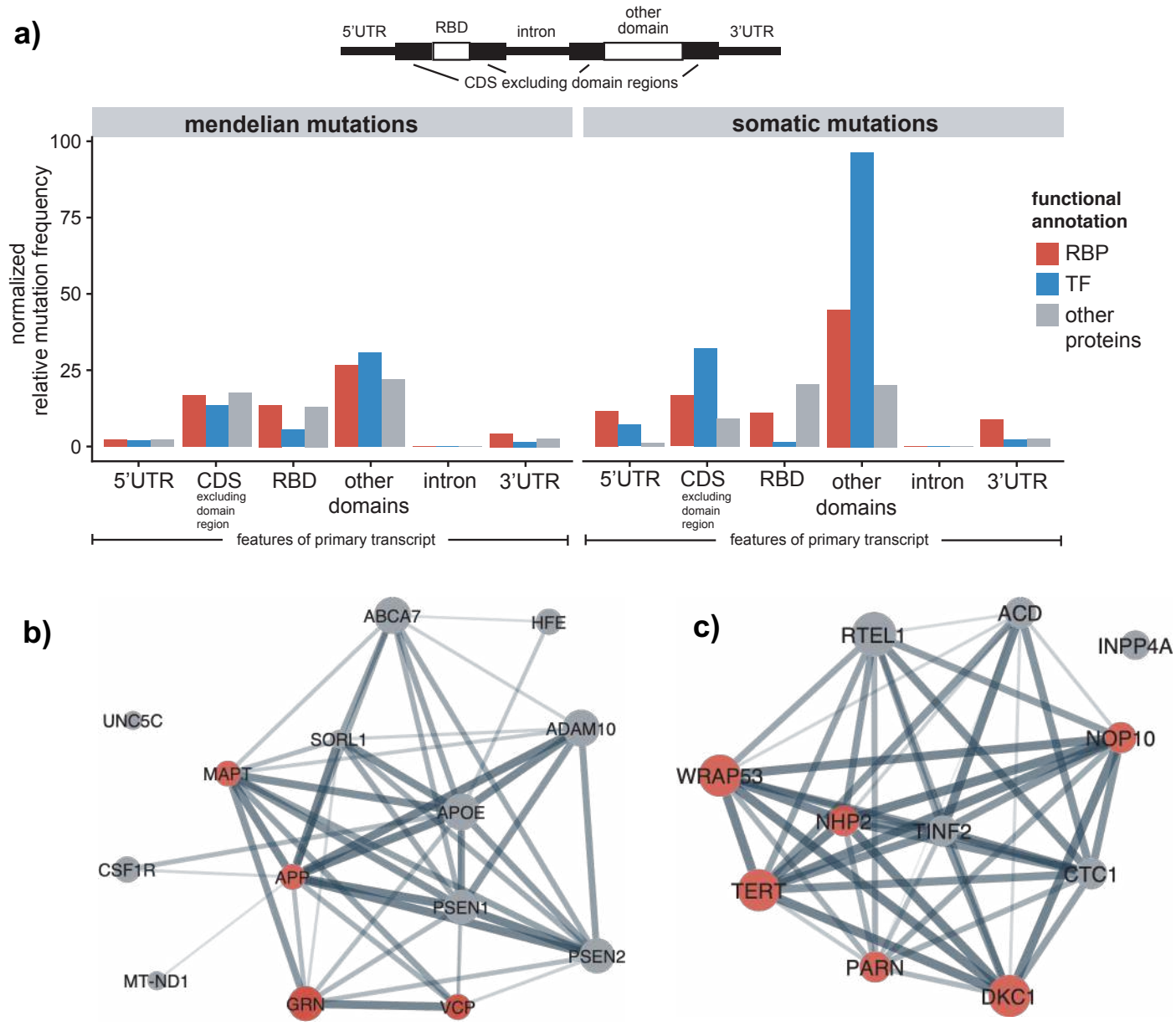

a)
5'UTR
RBD
intron
other domain
3'UTR
CDS excluding domain regions
mendelian mutations
somatic mutations
normalized relative mutation frequency
100
75
50
25
0
5'UTR
CDS excluding domain region
RBD
other domains
intron
3'UTR
features of primary transcript
functional annotation
RBP
TF
other proteins
b)
ABCA7
HFE
UNC5C
SORL1
ADAM10
MAPT
APOE
CSF1R
APP
PSEN1
PSEN2
MT-ND1
GRN
VCP
c)
RTEL1
ACD
INPP4A
NOP10
WRAP53
NHP2
TINF2
CTC1
TERT
PARN
DKC1

Figure 6

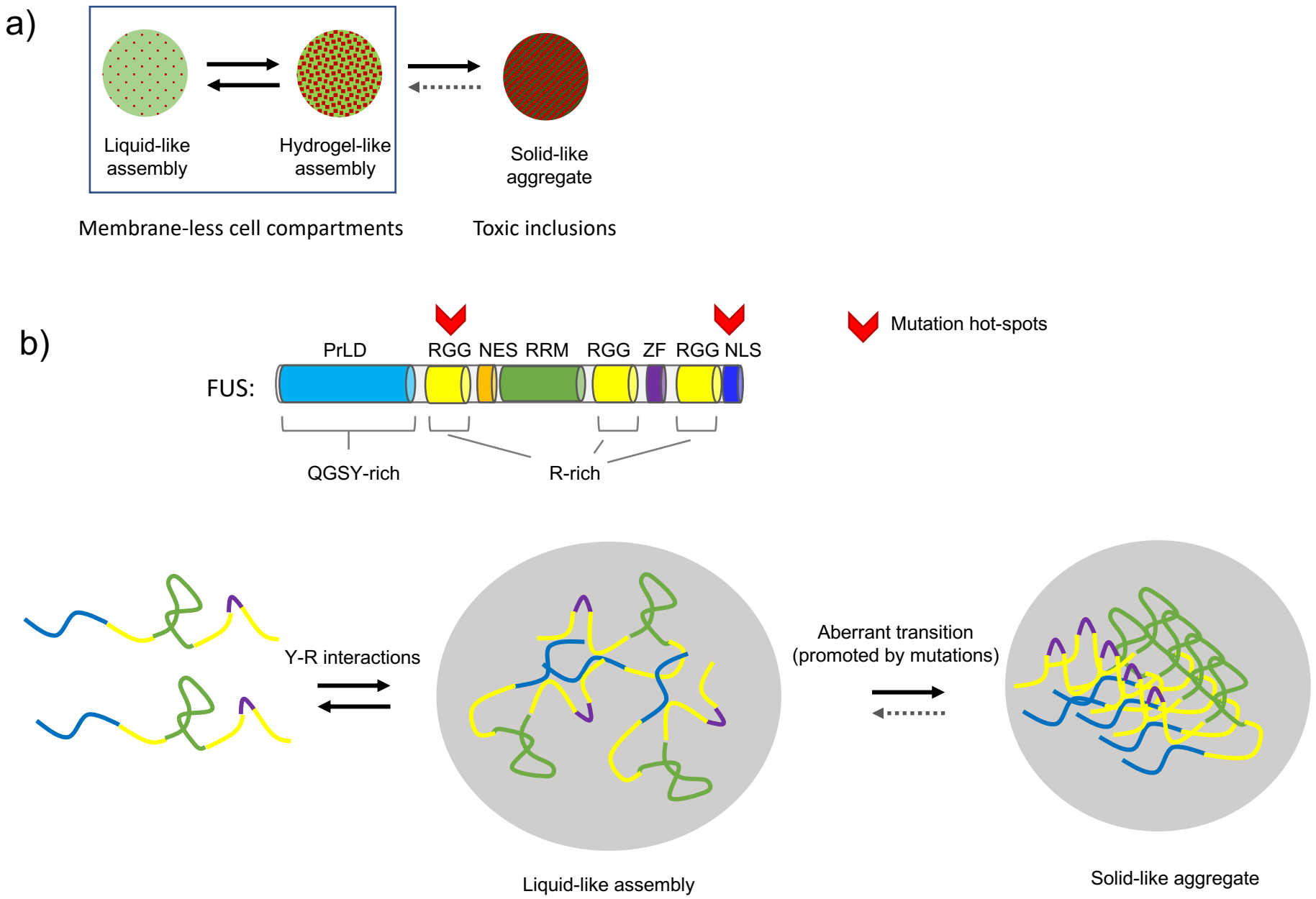

a)
Liquid-like assembly
Hydrogel-like assembly
Solid-like aggregate
Membrane-less cell compartments
Toxic inclusions
b)
Mutation hot-spots
FUS:
PrLD
RGG
NES
RRM
RGG
ZF
RGG
NLS
QGSY-rich
R-rich
Y-R interactions
Liquid-like assembly
Aberrant transition
(promoted by mutations)
Solid-like aggregate

Figure 7

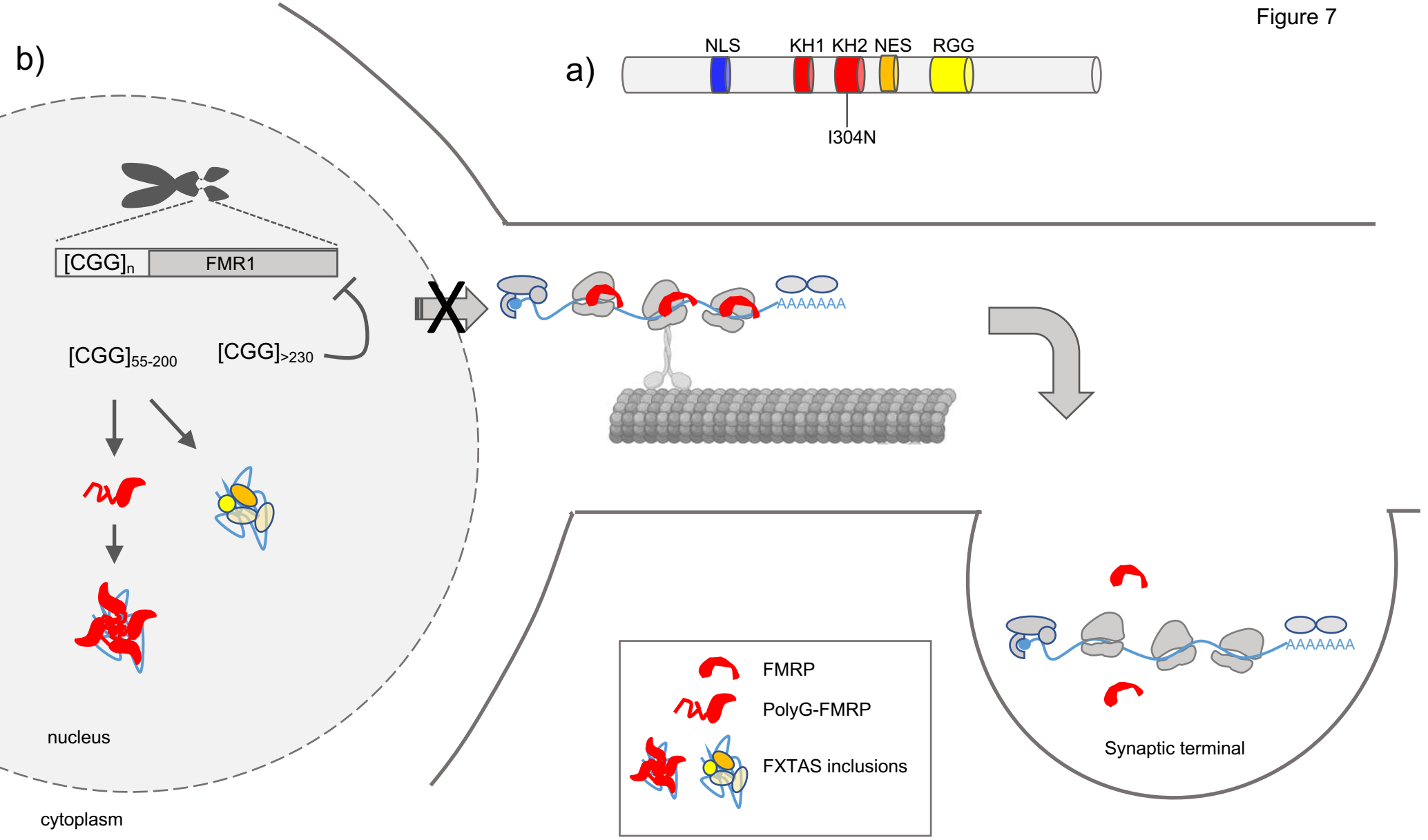
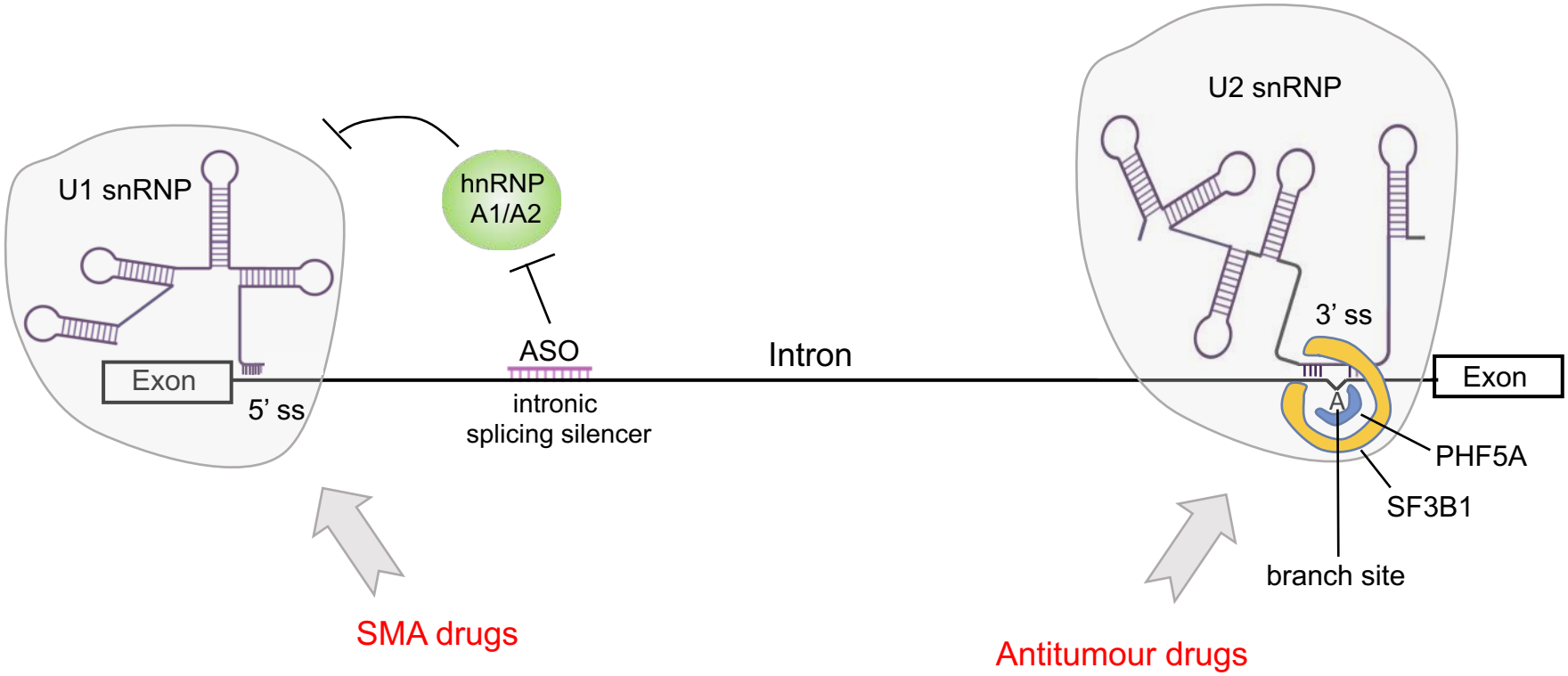

b)
a)
NLS
KH1
KH2
NES
RGG
I304N
[CGG]n
FMR1
[CGG]55-200
[CGG]>230
AAAAAAA
nucleus
cytoplasm
FMRP
PolyG-FMRP
FXTAS inclusions
AAAAAAA
Synaptic terminal

Figure Box 1


U1 snRNP
hnRNP
A1/A2
Exon
5' ss
ASO
intronic
splicing silencer
Intron
U2 snRNP
3' ss
Exon
A
PHF5A
SF3B1
branch site
SMA drugs
Antitumour drugs