

REVIEW

Open Access



The new (dis)order in RNA regulation

Aino I. Järvelin, Marko Noerenberg, Ilan Davis and Alfredo Castello*

Abstract

RNA-binding proteins play a key role in the regulation of all aspects of RNA metabolism, from the synthesis of RNA to its decay. Protein-RNA interactions have been thought to be mostly mediated by canonical RNA-binding domains that form stable secondary and tertiary structures. However, a number of pioneering studies over the past decades, together with recent proteome-wide data, have challenged this view, revealing surprising roles for intrinsically disordered protein regions in RNA binding. Here, we discuss how disordered protein regions can mediate protein-RNA interactions, conceptually grouping these regions into RS-rich, RG-rich, and other basic sequences, that can mediate both specific and non-specific interactions with RNA. Disordered regions can also influence RNA metabolism through protein aggregation and hydrogel formation. Importantly, protein-RNA interactions mediated by disordered regions can influence nearly all aspects of co- and post-transcriptional RNA processes and, consequently, their disruption can cause disease. Despite growing interest in disordered protein regions and their roles in RNA biology, their mechanisms of binding, regulation, and physiological consequences remain poorly understood. In the coming years, the study of these unorthodox interactions will yield important insights into RNA regulation in cellular homeostasis and disease.

Keywords: RNA-binding protein, Intrinsically disordered protein, Co- and post-transcriptional RNA regulation, RS repeat, RGG-box, GAR repeat, Basic patch, Poly-K patch, Arginine-rich motif, RNA granule

Plain English summary

DNA is well known as the molecule that stores genetic information. RNA, a close chemical cousin of DNA, acts as a molecular messenger to execute a set of genetic instructions (genes) encoded in the DNA, which come to life when genes are activated. First, the genetic information stored in DNA has to be copied, or transcribed, into RNA in the cell nucleus and then the information contained in RNA must be interpreted in the cytoplasm to build proteins through a process known as translation. Rather than being a simple process, the path from transcription to translation entails many steps of regulation that make crucial contributions to accurate gene control. This regulation is in large part orchestrated by proteins that bind to RNA and alter its localisation, structure, stability, and translational efficiency. The current paradigm of RNA-binding protein function is that they contain regions, or domains, that fold tightly into an ordered interaction platform that specifies how and where the interaction with RNA will occur. In this

review, we describe how this paradigm has been challenged by studies showing that other, hitherto neglected regions in RNA-binding proteins, which in spite of being intrinsically disordered, can play key functional roles in protein-RNA interactions. Proteins harbouring such disordered regions are involved in virtually every step of RNA regulation and, in some instances, have been implicated in disease. Based on exciting recent discoveries that indicate their unexpectedly pervasive role in RNA binding, we propose that the systematic study of disordered regions within RNA-binding proteins will shed light on poorly understood aspects of RNA biology and their implications in health and disease.

Background

Structural requirements for RNA-protein interactions

RNA-binding proteins (RBPs) assemble with RNA into dynamic ribonucleoprotein (RNP) complexes that mediate all aspects of RNA metabolism [1, 2]. Due to the prominent role that RBPs play in RNA biology, it is not surprising that mutations in these proteins cause major diseases, in particular neurological disorders, muscular atrophies and cancer [3–7]. Until recently, our understanding of how

* Correspondence: alfredo.castellopalomares@bioch.ox.ac.uk
Department of Biochemistry, University of Oxford, South Parks Road, Oxford
OX1 3QU, UK

RBPs interact with RNA was based on a limited number of globular RNA-binding domains (RBDs), which include RNA-recognition motif (RRM), K-homology domain (KH), double-stranded RBD (dsRBD), zinc fingers (Znf), DEAD box helicase domain, and others (for recent reviews, see [8–10]). Each of these RBDs interacts with RNA following distinct mechanisms and differ in specificity and affinity for their target RNA. Promiscuous RNA binding is often mediated by interactions with the phosphate-sugar backbone, whereas sequence-specificity builds on interactions with the nucleotide base and shape complementarity between protein and RNA interfaces. While the most common RBDs interact with short (4–8 nt) sequences, others display lower or complete lack of sequence selectivity, recognising either the RNA molecule itself or secondary and three-dimensional structures [8, 11]. As the affinity and specificity of a single RBD is often insufficient to provide selective binding *in vivo*, RBPs typically have a modular architecture containing multiple RNA-interacting regions [8]. RNA-binding proteins are typically conserved, abundant, and ubiquitously expressed, reflecting the core importance of RNA metabolism in cellular physiology [12, 13].

The coming of age for RNA-binding proteins — the emerging role of protein disorder

Early on, it was recognised that not all RNA-binding activities could be attributed to classical RBDs. Computational predictions based on transcriptome complexity suggested that 3–11 % of a given proteome should be dedicated to RNA binding, whereas only a fraction of this number could be identified by homology-based searches for classical RBDs [14, 15]. Moreover, there were several reports of RNA-binding activities within protein domains lacking similarities to any classical RBD [16, 17]. A number of studies showed that intrinsically disordered regions, lacking any stable tertiary structure in their native state, could contribute to RNA binding. For example, the flexible linker regions that separate the two RRM of the poly(A)-binding protein (PABP) and polypyrimidine tract binding protein 1 (PTBP1), not only orientate the domains with respect each other, but also mediate RNA binding [18]. Flexible regions in RBPs rich in serine and arginine (S/R) and arginine and glycine (R/G) were found to contribute, or even to account for, RNA-binding activities [19, 20]. Furthermore, early computational analyses revealed that proteins involved in transcription and RNAs processing are enriched in disordered protein regions [21, 22], hinting on a broader role for protein disorder in RNA metabolism.

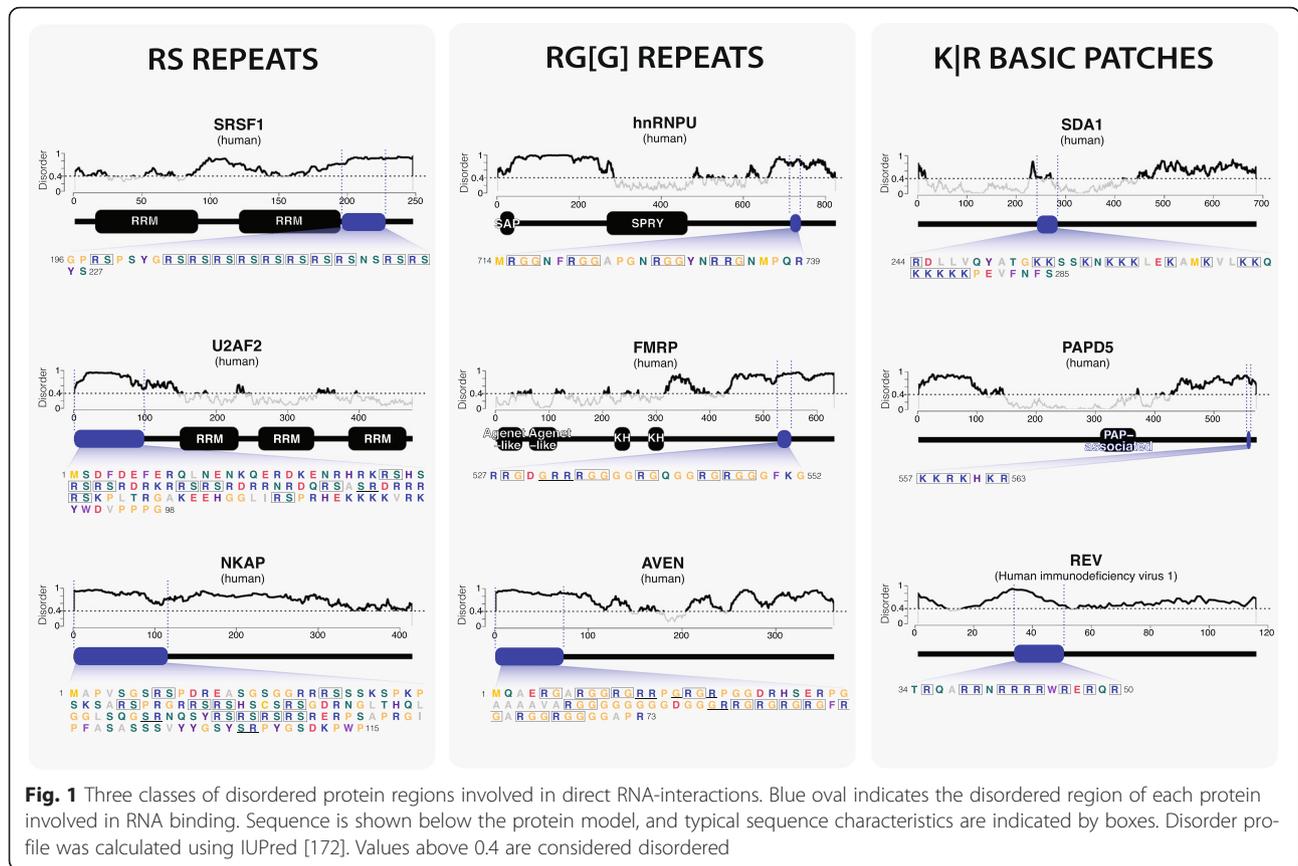
Recently, the development of proteome-wide approaches for comprehensive determination of the RBP repertoire within the cell (RBPome) has substantially increased the number of known unorthodox RBPs. *In vitro*

studies in yeast identified dozens of proteins lacking classical RBDs as putative RBPs, including metabolic enzymes and DNA-binding proteins [23, 24]. Two recent studies that employed *in vivo* UV crosslinking, poly(A)-RNA capture, and mass spectrometry, identified more than a thousand proteins interacting with RNA, discovering hundreds of novel RBPs [25, 26]. Strikingly, both known and novel RBPs were significantly enriched in disordered regions compared with the total human proteome. Approximately 20 % of the identified mammalian RBPs (~170 proteins) were disordered by over 80 % [25, 27]. Apart from the disorder-promoting amino acids such as serine (S), glycine (G), and proline (P), these disordered regions were enriched in positively (K,R) and negatively (D, E) charged residues as well as tyrosine (Y) [25], amino acids often found at RNA-interacting surfaces in classical RBDs [8]. Disordered amino acid sequences in RBPs form recognisable patterns that include previously reported motifs such as RG- and RS-repeats as well as new kinds of motifs, such as K or R-rich basic patches (Fig. 1). As with classical RBDs, disordered regions also occur in a modular manner in RBPs, repeating multiple times in a non-random manner across a given protein and, in some instances, combining with globular domains [25]. Taken together, these observations suggest that disordered regions 1) contribute to RBP function; 2) combine in a modular manner with classical RBDs suggesting functional cooperation; and 3) may play diverse biological roles, including RNA binding. Supporting this, a recent report has shown that globular RBDs are on average well conserved in number and sequence across evolution, while disordered regions of RBPs have expanded correlating with the increased complexity of transcriptomes [13]. What is the contribution and functional significance of protein disorder in RNA-protein interactions? Below, we will discuss what is known about disordered regions in RNA binding and metabolism, as well as physiology and disease, based on accumulating literature (Table 1, Additional file 1: Figure S1).

Review

Disordered RS repeats put RNA splicing in order

Disordered, arginine and serine (RS) repeat containing regions occur in a number of human proteins referred to as SR proteins and SR-like proteins (reviewed in [28, 29]). SR proteins are best known for their roles in enhancing splicing but have been ascribed functions in other RNA processes from export, translation, and stability to maintenance of genome stability (e.g. [30, 31] for reviews). There are twelve SR proteins in human that contain 1–2 classical RRMs and an RS repetitive motif of varying length [30]. Classical SR proteins bind exonic splicing enhancers in nascent RNA through their RRM



and promote splicing of adjacent introns [32, 33]. The RS repeat enhances splicing in a length-dependent manner [34]. RS repeats are predicted to be intrinsically disordered [35] (Table 1), but phosphorylation promotes a transition towards a less flexible, arch-like structure with an influence on RNA binding in the serine/arginine-rich splicing factor 1 (SRSF1) [36] (Fig. 1). RS repeats have been shown to directly bind RNA during multiple steps of splicing [19, 37–39] and to contribute to binding affinity of RRM for RNA by inducing a higher affinity form of the RRM [40]. RS repeats can also mediate protein-protein interactions [28, 33], hence their association with RNA can also be indirect. RS-mediated protein binding seems to be compatible with RNA binding [33, 41], suggesting that protein and RNA binding could take place simultaneously or sequentially. RNA-binding by RS repeats seems to be rather non-specific, as motif shortening, replacement of arginine for lysine, amino acid insertion, and replacement for a homologous sequences are well tolerated [19, 37, 38]. In summary, there is compelling evidence that disordered RS protein motifs play an important role in RNA splicing, and that the interaction between these repeats and RNA occurs mostly in a sequence-independent manner. Nevertheless, it remains

to be determined how many of the SR proteins interact with RNA through the RS repeats, and whether the differences in RS repeat length have a direct effect on RNA binding affinity or specificity.

Certain members of the SR-related protein family lack RRM and are involved in diverse RNA metabolic processes [42]. For example, NF-kappa-B-activating protein (NKAP) (Fig. 1) is an SR-related protein, with a newly discovered role in RNA splicing [43], but originally known for its roles in NF-kappa-B activation [44] and as a transcriptional repressor of Notch-signalling in T-cell development [45]. This protein binds RNA through its RS repeat, in cooperation with an RBD at its C-terminal region. A transcriptome-wide study showed this protein targets diverse classes of RNAs, including pre-mRNAs, ribosomal RNAs and small nuclear RNAs [43]. RNA-binding RS repeat sequences can also be found in viral proteins, such as the nucleocapsid of severe acute respiratory syndrome coronavirus (SARS-CoV), causative agent of the alike-named disease. This protein employs RS-rich disordered region, in cooperation with other RNA-binding regions, to capture viral RNA and package it into virions [46]. Taken together, these reports suggest that RS repeats have broader roles in RNA-binding than previously anticipated.

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172]

Protein						Properties of disorder involved in RNA binding						
ID	Name	Aliases	Species	Canonical domains	Function	Class	Sequence	Disorder assignment	Target RNA preference	Regulation at disordered region	Interaction with other biomolecules	Ref
SRSF1	Serine/arginine-rich splicing factor 1	ASF, SF2, SF2P33, SFRS1	Homo sapiens	2xRRM	RNA splicing. Essential for heart development.	RS	¹⁹⁶ -GPRSPSYGRSRSRSR SRSRSRSRSNSRSRS YS ⁻²²⁷	Experimental	-	Serine phosphorylated. Becomes more structured upon phosphorylation. Alternatively spliced.	Protein	[36, 39, 173–175]
U2AF2	Splicing factor U2AF 65 kDa subunit	U2AF65	Homo sapiens	3xRRM	RNA splicing.	RS	¹ -MSDFDEFERQLNENK QERDKENRHRKRSHS RSRSRDRKRRSRSRD RRNRDQRSASRD RSKPLTRGAKEEHGG LIRSPRHEKKKVRK YWDVPPPG ⁻⁹⁸	Predicted	No specificity	Serine phosphorylation, lysine acetylation, lysine hydroxylation ^a	Protein	[19, 176]
NKAP	NF-kappa-B-activating protein	-	Homo sapiens	None	RNA splicing, transcriptional repression.	RS	¹ -MAPVSGSRSPDREAS GSGGRRSSSKSPKP SKSARSPRGRSRSH SCSRSGDRNGLTHQL GGLSQGSRNQSYRSR SRSRSRERPSAPRGI PFASASSVYGYYS RPYGSDKPWP ⁻¹¹⁵	Predicted	poly (U)	Lysine acetylation ^a	Protein	[43]
Nucleo-capsid protein	-	Nucleoprotein, NC, N	Severe acute respiratory syndrome coronavirus (SARS-CoV)	None	Major structural component of virions that associates with genomic RNA to form a long, flexible, helical nucleo-capsid.	Other, RS, polyK/ other	¹ -MSDNGPQSNQRSAPR ITFGGPTDSTDNNQN GGRNGARPKQRRPQ ⁻⁴⁴ , ¹⁸² -QASSRSSRSRGNR NSTPGSSRGNSPARM ASGGGETALALLLDR LNQLESKVSQGGQQQ QGQTV ⁻²⁴⁷ , ³⁶⁶ -PTEPKKDKKKKTDEA QPLPQRQKKQPTVTL LPAADMDDFSRQLQN SMSGASADSTQ ⁻⁴²²	Experimental	poly (U) ssRNA	-	-	[46, 177]
ALYREF	Aly/REF export factor 2	Alyref	Mus musculus	1xRRM	RNA export.	RG	²² -VNRGGGPRRNRPAIA RGGRNRPAYSR ⁻⁴⁸	Experimental	-	TAP displaces RNA from ALYREF	Protein	[54, 55, 57, 178]
Aven	Cell death regulator Aven	-	Homo sapiens	None	Positive translational regulator.	RG	¹ -MQAERGARGGRRRP GRGRPGGDRHSERPG AAAAVARGGGGGGG DGGRRRGRGRGRFR GARGRGGGAPR ⁻⁷³	Predicted	RNA G-quadruplex	Methylated (no influence on RNA binding; influences protein interactions and polysome association). Alternative transcript (mouse)	Protein	[179, 180]
Caprin-1	-	-	-	None	-	RG	-	Predicted	-	-	-	-

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

		GPIAP1, GPIIP137, M11S1, RNG105	Homo sapiens, Xenopus		Regulation of localised translation, synaptic plasticity, cell proliferation and migration.	RG	⁶¹² RGGSRGARGLMNGYR GPANGFRGGYDGYRP SFNNTPNISGYTQSQF SAPRDYSGYQRDGYQ QNFKRGSQSGPRGA PRGRGGPPRPNRGMP QMNTQQV ⁻⁷⁰⁸ (human), ⁵⁷⁸ RGMARGGQRGNRGM Ngyrgqsngfrgg ⁻⁶⁰⁵ (Xenopus)			The end of the human sequence (RGGPPRP NRGMPQMNTQQV) is in an alternative isoform ^a		[181, 182]
DDX4	Probable ATP-dependent RNA helicase DDX4	Vasa	Homo sapiens	None	RNA helicase.	RG	¹ MGDEWAEINPHIMS SYVPIFEKDRYSGEN GDNFNRTPASSEMID DGPSRRDHFMSGFA SGRNFGNRDAGECNK RDNTSTMGGFVGS FGNRGFSNSRFEDGD SSGFWRESSNDCEDN PTRNRGFSKRGGYRD GNNSEASGPYRRGGR GSFRGCRGGFLGSP NNDLDPDECMQRTGG LFGSRRPVLSTGNG DTSQSRSGSGSERGG YKGLNEEVITGSGKN SWKSEAEAGES ⁻²³⁶	Experimental	Single-stranded DNA.	Arginine methylation. Alternative isoforms ^a	-	[130]
EWS	RNA-binding protein EWS	EWSR1	Homo sapiens	1xRRM	Transcription, splicing.	RG	²⁸⁸ PGENRSMSPDNRGR GRGGFDRGGMSRGG GGGRGGMGSAGERGG FNKPGGPMDEGPDLD LGPPVDp ⁻³⁵⁴ , ⁴⁵⁰ PMNSMRGGLPPREGR GMPPLRGPGGGPGG PGGPMGRMGRRGGDR GGFPPRG ⁻⁵⁰¹ , ⁵⁴⁵ APKPEGFLPPFP GGDRGRGGPGGMRGG RGGLMDRGGPGGMFR GGRGGDRGGFRGG MDRGGFGGRRGGPG GPPGPLMEQMGGRRG GRGGPGKMDK GEHRQERRDRPY ⁻⁶⁵⁶	Predicted	G-quadruplex (RGG3, not RGG1 or RGG2)	Alternative splicing ^a . Arginine dimethylation at RGG repeats affects protein sub cellular localization	DNA (via RGG3). All three RGG repeats bind SMN protein.	[183-187]
FMRP	Fragile X mental retardation protein 1	FMR1	Homo sapiens, mouse	2xKH	Regulation of translation (repressor).	RG	⁵²⁷ RRGDGRRRGGGGRGQ GGRGRGGGFKG ⁻⁵⁵²	Experimental	G quartets, G-quadruplex	Arg methylation. Alternative splicing at regions flanking the RGG-box alters FMRP's capacity to bind RNA, to be methylated, and	C-terminal part of this protein that also includes the RG region is involved in protein-	[68-70, 72, 75-78, 152,

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

FUS	RNA-binding protein FUS	TLS	Homo sapiens, Drosophila melanogaster	1xRRM	Splicing, poly-adenylation.	RG	213 ⁻ RGGRGRGG ⁻²²⁰ , 241 ⁻ PRGRGGGRGRGG ⁻²⁵³ , 377 ⁻ RGGGNGRGGGRGGP MGRGGYGGGGSGGGG RGG ⁻⁴⁰⁹ , 472 ⁻ RRGGRRGGYDRGGYRG RGDRGGFRGGRRGG DRGG ⁻⁵⁰⁵	Predicted	G-quadruplex	associate with polysomes. Arginine methylation.	protein interactions.	188, 189
hnRNP U	Heterogeneous nuclear ribonucleoprotein U	HNRPU, SAFA, U21.1	Homo sapiens	None	RNA stability, U2 snRNP maturation, DNA binding.	RG	714 ⁻ MRGGNFRGGAPGNRG GYNRRGNMPQR ⁻⁷³⁹	Predicted	Poly (U and poly (G) homopolymers, UGUGG	-	DNA	[20, 51]
ICP27	Infected cell protein 27, Immediate-early protein IE63	-	Herpes simplex virus	None	RNA export.	RG	138 ⁻ RGRRRRRRGRGRGG ⁻¹⁵²	Predicted	poly (G) and poly (U) homopolymers, GC-rich sequences	Methylated	-	[194–196]
LAF1	-	DDX3	C. elegans	None	RNA helicase.	RG	1 ⁻ MESNQSNNGSGNAA LNRGGRYVPPHLRGG DGGAAAAASAGDDDR RGGAGGGGYRRGGGN SGGGGGGYDRGYND NRDDRNRGGSGGYG RDRNYEDRGYNGGGG GGGNRGYNNRGGGG GGYNRQDRGDGGSSN FSRGGYNNRDEGSDN RGSGRSYNNDRRDNG GDG ⁻¹⁶⁸	Experimental	-	Region 43–106 containing RG-repeat is alternative.	-	[142]
NXF1	Nuclear RNA export factor 1	TAP	Mus musculus, homo sapiens	None	Nuclear export.	RG	2 ⁻ ADEGKSYSEHDDERV NFPQRKKKGRGPPFRW KYGEGNRRSRGGSG IRSSRLEEDDGDVAM SDAQDQPRVRYNPYT TRPNRRGDTWHDRDR IHVTVRRDRAPPERG GAGTSQDGTSKN ⁻¹¹⁸	Predicted	Non-specific	-	Protein. Overlaps a nuclear localisation and export signals.	[55, 197, 198]
Nucleolin	-	NCL, Protein C23	Hamster	4xRRM	Chromatin decondensation, pre-rRNA transcription, ribosome assembly.	RG	630 ⁻ MEDGEIDGNKVTLDW AKPKGEGFGGRGGG RGGFGGRGGGRGGGR GGFGGRGRGGFGGRG GFRGGRGGGGGGDF KPQGGKTKFE ⁻⁷¹⁴	Experimental. Suggested to form a flexible β-spiral.	None	-	Protein (in human)	[199, 200]
RBMX	RNA-binding motif protein, X chromosome	HNRPG, RBMXP1	Homo sapiens, Xenopus laevis	1xRRM	Regulation of transcription, splicing.	RG	333 ⁻ DLYSSGRDRVGRQER GLPPSMERGYPPPRD SYSSSRGAPRGGGR GGSRSRDRGGRSR ⁻³⁹⁰	Predicted	C-terminal regions binds structured (hairpin) RNA	Identical C-terminal sequence is mouse RBMX is alternatively spliced.	-	[201–206]

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

Foamy virus Gag	-	-	Human foamy virus	None	Viral genome binding, capsid formation.	RG	⁴⁸⁵ -RPSRGRGRGQN ⁻⁴⁹⁵	Predicted	-	-	-	[207–210]
TERF2	Telomeric repeat-binding factor 2	TRBF2, TRF2	Homo sapiens	None	Presynaptic plasticity, axonal mRNA transport, telomere maintenance	RG	⁴³ -MAGGGGSSDGSGRAAGRRASRSSGRARRGRHEPGLGGPAERGAG ⁻⁸⁶	Predicted	G-rich, TERRA	Arginine methylation	Protein	[211–214]
XTUT7	-	-	Xenopus laevis	Zinc finger	RNA polyuridylation, translational repression.	Basic patch (poly R)	⁴⁵³ -MRRNRVRRRNENAGNQR ⁻⁴⁷¹	Predicted	-	-	-	[215]
Tat	Transactivating regulatory protein	-	Human immunodeficiency virus (HIV)	None	transcriptional activator, transcription elongation.	Basic patch (poly R)	⁴⁹ -RKKRRQRRR ⁻⁵⁷	Experimental	Structured RNA (HIV-1 Trans-activation response element, TAR)	Arginine methylation (with impact on RNA binding). Lysine acetylation (impact on TAR binding, through an effect on Tat-TAR-CyclinT1 ternary complex formation).	Protein	[85, 88–91, 93, 216–223]
Rev	Regulator of expression of viral proteins	-	Human immunodeficiency virus (HIV)	None	RNA export.	Basic patch (poly R)	³⁴ -TRQARRNRRRRWRERQR ⁻⁵⁰	Experimental	Structured RNA (HIV-1 Rev response element, RRE)	Arginine methylation.	Protein	[96–101, 103, 104, 153, 154, 224]
Tat	Transactivating regulatory protein	S ORF, bTat	Bovine immunodeficiency virus	None	Transcriptional activator	Basic patch (polyR)	⁷⁰ -RGTRGKRRIR ⁻⁸¹	Experimental	Structured RNA (TAR)	-	Protein	[91]
Coat protein	-	-	Alfalfa mosaic virus	None	Capsid protein, viral RNA. Translation initiation.	Basic patch (poly K)	⁶ -KKAGGKAGKPTKRSQNYAALRK ⁻²⁷	Experimental	-	-	-	[225, 226]
PAPD5	Non-canonical poly (A) RNA polymerase PAPD5	-	Homo sapiens	None	RNA oligoadenylation, RNA stability	Basic patch (poly K)	⁵⁵⁷ -KKRKHKR ⁻⁵⁶³	Predicted	May have a preference for structured RNA	Alternative splicing ^a	-	[109]
SDAD1	Protein SDA1 homolog	-	Homo sapiens	None	Protein transport, ribosomal large subunit export from nucleus.	Basic patch (poly K)	²⁴⁴ -RDLLVQYATGKKSSKNKKKLEKAMKVLKKQKKKKKPEVFNFS ⁻²⁸⁵	Predicted	-	-	-	[58]
HMGA1	-	-	-	None	-	(e) AT	²¹ -TEKRGRPRK ⁻³¹	Experimental	-	-	DNA	

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

	High mobility group protein HMG-I/HMG-Y		Homo sapiens						Binds structured RNA.	Arginine methylation.		[121, 124, 125, 127]
Tip5	Bromodomain adjacent to zinc finger domain protein 2A	BAZ2A	Homo sapiens	None	Epigenetic rRNA gene silencing.	(e) AT	⁶⁵⁰ GKRGRPRNTEK ⁻⁶⁶⁰ , ⁶⁷⁰ KRGRGRPPKVKIT ⁻⁶⁸²	Experimental	Exhibits preferential binding towards dsRNA	-	DNA	[127, 227, 228]
PTOV1	Prostate tumor-overexpressed gene 1 protein	ACID2, PP642	Homo sapiens	None	Regulation of transcription.	(e) AT	¹ MVRPRRAPYRSGAGG PLGGRRPRRPLVVR AVRSRSWPASPRG ⁻⁴³	Predicted	Exhibits preferential binding towards dsRNA	Alternative splicing ^a	DNA	[127]
GPBP1	-	Vasculin, GPBP, SSH6	Homo sapiens	None	Transcription factor, positive regulation of transcription	e (AT)	³⁸ NRVDVNRHSSDGG FDSAIGRPNNGNFGFR KEKNGWRTHGRNG ⁻⁸⁰	Predicted	Exhibits preferential binding towards dsRNA	Alternative splicing ^a	DNA	[127]
SRSF2	Serine/arginine-rich splicing factor 2	SFRS2	Homo sapiens	1xRRM	RNA splicing.	Other (GRP)	¹ MSYGRPPP ⁻⁸ , ⁹³ GRPPDSHHS ⁻¹⁰¹	Experimental	UCCA/UG, UGGA/UG	-		[229, 230]
Tra2-β1	Transformer-2 protein homolog beta	TRA2B, SFRS10	Homo sapiens	1xRRM	RNA splicing.	Other	¹¹⁰ NRANPDPNCC ⁻¹¹⁹ , ¹⁹⁴ SITKRPH ⁻²⁰¹	Experimental	GAAGAA (primary), AGAAG (primary), GACUACAACA AGUC (structured)	-		[40, 231–233]
hnRNP A1	Heterogeneous nuclear ribonucleoprotein A1	HNRPA1	Human, Xenopus tropical	2xRRM	hnRNP particle formation, nucleocytoplasmic transport, splicing.	Other/ RG	¹⁸⁶ MASASSQRGRSGSG NFGGGRRGGFGGNDN FGRGGNFSGRGGFGG SRGGGGYGGSGDGYN GFGNDGGYGGGGPGY SGGSRGYGGGQGYG NQGGYGGSGSYDSY NNGGGGGFGGSGSN FGGGGSYNDFGNYYN QSSNFGPMKGGNFGG RSSGPGYGGGGQYFAK PRNQGGYGGSSSSS YGSRRF ⁻³⁷²	Predicted	-	Region containing the RG- and FG-repeat peptides is alternatively spliced. RG-region may mediate RNA binding. The entire region is involved in hnRNP A1 aggregation and includes a nuclear targeting sequence.	-	[136, 234–237]
LUZP4	Leucine zipper protein 4	CT-28,	Homo sapiens	None	Nuclear export.		⁵¹ RQNHSSKESPSRQQSKAH RHRHRRGYSRCR ⁻⁸⁰ , ²³⁸ LVD TQSDLIATQRDLIATQK DLIATQRDLIATQRDLIVTQR DLVATERDL ⁻²⁸⁷	Predicted	-	Alternative splicing affecting the first, R-rich region ^a	Protein	[197]
ORF57	-			None		Other		Experimental			Protein	[178]

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

	52 kDa immediate-early phosphoprotein, mRNA export factor ICP27 homolog		Herpes-virus saimiri		Viral RNA regulation.		64-RQRSPITWEHQSPLS RVYRSPSPMRFQKRP RISSNSTSRSCKTSW ADRVREAAAQR ¹²⁰		Viral RNA: GAAGAGG, CAGUCGCG AAGAGG	RNA binding region partially overlaps with ALYREF binding site.		
APC	Adenomatous polyposis coli protein	-	Mus musculus	None	Microtubule binding, negative regulator of Wnt signaling.	Other	2223-SISRGRTMIHIPGLR NSSSSTSPVSKKGGP LKTPAKSPSEGGGA TTSRPGTKPAGKSEL SPITRQTSQISGSNK GSSRSGSRDSTPSRP TQQPLSRPMQSPGRN SISPGRNGISPPNKL SQLPRTSSPSTASTK SSGSGKMSYTSRGRQ LSQQNLTKQASLSKN ASSIPRESASKGLN QMSNGNGSNKVELS RMSSTKSSGSESDSS ERPALVRQSTFIKEA PSPTLRRKLEESASF ESLSPSSRPDSPTRS QAQTPVLSPLPDMS LSTHPSVQAGGWRKL PPNLSPTIEYNDGRP TKRHDIARSHSESPS RIPINRAGTWKREHS KHSSSLPRVSTWRRT GSSSILSASSE ⁻²⁵⁷⁹	Predicted	G-rich motif	-	-	[238]
CTCF	Transcriptional repressor CTCF	-	Homo sapiens	11x Zn finger (3 according to Pfam)	-	Other	575-DNCAGPDGVEGENGG ETKKSQRGRKRKMRS KKEDSSDSENAEPDL DDNEDEEPPAVEIEP EPEPQPVTPAPPPAK KRRGRPPGRTNQPKQ NQPTAIQVEDQNTG AIENIIVEVKKEPDA EPAEGEEEEQAAT DAPNGDLTPEMILSM MDR ⁻⁷²⁷	Predicted	-	Serine phosphorylation ^a	-	[239]
Df31	Decondensation factor 31	Anon1A4	D. melanogaster	None	Regulation of higher-order chromatin structure, maintenance of open chromatin.	Other	1-MADVAEQKNETPWVE KVAAEEVDVAVKKDAV AAEEVAAEKASITEN GGAEESVAKENGAA DSSATEPTDAVDGEK ASEPTVSFAADKDEK KDEKEDSAADGED TKKESSEAVLPAVEN GSEEVINGDSTDAPA IEAVKRKVDEAAKA DEAVATPEKKAKLDE	Experimental	Non-specific but does not bind ssDNA or dsDNA. Preferentially binds snoRNA.	-	-	[127, 240]

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

Ezh2	Histone-lysine N-methyltransferase EZH2	Enx1h	Mus musculus	None	Polycomb group protein. Involved in H3 methylation (H3K9me and H3K27me).	Other	ASTKDEVQNGAEASE VAA ⁻¹⁸³ 342-RIKTPPKRPGGRRRG RLPNNSSRPSTPTI ⁻³⁷⁰	Predicted	May have a preference for RNA stem loops.	1st Thr is phosphorylated in a cell cycle dependent manner. Phosphorylation increases RNA binding.	This region overlaps a region involved in protein-protein interactions in human, however, RNA and protein binding regions may be distinct from one another.	[241-243]
Nrep	Neuronal regeneration-related protein	P311	Mus musculus	None	Axonal regeneration, cell differentiation.	Other	27-KGRLPVPKEVNRKKM EETGAASLTPPGSRE FTSP ⁻⁶⁰	Experimental	-	-	Protein	[244]
Gemin5	Gem-associated protein 5	-	Homo sapiens	None	snRNP assembly, splicing, IRES-mediated translation initiation.	Other	1297-PNSSVWRAGHRTL SVEPSQLDTASTEET DPETSQPEPNRPSSEL DLRLTEEGERMLSTF KELFSEKHASLQNSQ RTVAEVQETLAEMIR QHQKSQLCKSTANGP DKNEPEVEAEQ ^{-1412, 1383-} EMIRQHQSCKSTANG PDKNEPEVEAEQ PLCSSSQCKEENEP LSLPELTKRLTEANQRMAK FPESIKAWPPFDVLECCL VLLLIRSHFPGCLAQEMQ QQAQELLQKYGNTK TYRRHCQTFCM ⁻¹⁵⁰⁸	Experimental	-	-	-	[245]
Nup153	-	-	Homo sapiens	None	Component of the nucleopore, RNA trafficking.	Other	250-KTSQLGDSPPFYPGKT TYGGAAA AVRQSKLR NTPYQAPVRRQMKAK QLSAQSYGVTSSSTAR RILQSLEKMSSPLAD AKRIPSIVSSPLNSP LDRSGIDITDFQAKR EKVDSQYPPVQRLMT PKPVSIATNRSVYFK PSLTPSGEFRKTNQR I ⁻⁴⁰⁰	Predicted	Single-stranded RNA with little sequence preference	Serine and threonine phosphorylation ^a	-	[246, 247]
SCML2	Sex comb on midleg-like protein 2	-	Homo sapiens	None	Binds Polycomb Repressive	Other	256-SPSEASQHSMSQSPQK TTLILPTQQVRRSSR IKPPGPTAVPKRSSS	Predicted	No specificity, but discriminates	Alternative isoform ^a ,	-	[248]

Table 1 Examples of RNA binding proteins where a disordered, non-classical region is involved in direct RNA binding. Additional details for each protein are presented in Additional file 1: Figure S1. Disorder prediction was calculated using IUPred [172] (Continued)

					Complex 1 and histones. Involved in epigenetic silencing.		VKNITPRKKGPNSGK KEKPLPVICSTSAAS ⁻³³⁰		between RNA and DNA.	Serine phosphorylation ^a			
KDM4D	Lysine-specific demethylase 4D	JMJD2D	Homo sapiens	None	Demethylates lysine 9 on histone H3.	Other	³⁴⁸ MEPRVPASQELSTQK EVQLPRAALGLRQL PSHWARHSPWPMAAR SGTRCHTLVCSLPR RSAVSGTATQPRAAA VHSSKKPSSTPSSTP GPSAQIIHPSNGRRG RGRPPQKLRAQELTL QTPAKRPLLAGTTCT ASGPEPELPELGAL MDKPVPLSPGLQHPV KASGCSWAPVP ⁻⁵²³	Experimental	-	-	-	[249]	
-	-	-	Synthetic	None	Bind HIV RNA (RRE)	Other/ polyR	SRSSRRNRNRNRNRNR, NHRRRRRQRNRNRNR, SPCRSRRSGSSRRNRNRNR	Experimental	Structured RNA (HIV-1 Rev response element, RRE)	-	-	-	[105]

^a According to uniprot, from a large-scale study but no detailed experimental confirmation available

RG-rich repeats — The swiss-army knife of protein-RNA interactions

A commonly occurring disordered RNA-binding motif in RBPs consists of repeats of arginine and glycine, termed RGG-boxes or GAR repeats. These sequences are heterogeneous both in number of repeats and in their spacing. A recent analysis divided these RG-rich regions into di- and tri-RG and -RGG boxes, and identified instances of such repeats in order of tens (di- and tri-RGG) to hundreds (tri-RG) and nearly two thousand (di-RG) proteins [47]. Proteins containing such repeats are enriched in RNA metabolic functions [47]. However, it is not currently clear whether the different repeat architectures provide distinct functional signatures.

The RGG box was first identified in the heterogeneous nuclear ribonucleoprotein protein U (hnRNP-U, also known as SAF-A) as a region sufficient and required for RNA binding (Table 1, Fig. 1). hnRNP-U lacks canonical RBDs, but has semi-structured SAP domain involved in DNA binding [48–50]. hnRNP-U has been found to target hundreds of non-coding RNAs, including small nuclear (sn)RNAs involved in RNA splicing, and a number of long non-coding (lnc)RNAs, in an RGG-box-dependent manner [51]. RGG-mediated interaction of hnRNP-U with the lncRNAs Xist [52] and PANDA [53] has been implicated in epigenetic regulation.

RG[G]-mediated RNA binding also plays a role in nuclear RNA export, as illustrated by the nuclear RNA export factor 1 (NXF1). While NXF1 harbours an RRM capable of binding RNA [54], most of the *in vivo* RNA-binding capacity is attributed to the RGG-containing, N-terminal region [55] (Table 1). The arginines in this motif play a key role in the interaction with RNA, which has been shown to be sequence-independent but necessary for RNA export [55]. NXF1 overall affinity for RNA is low [55, 56], and requires the cooperation with the export adapter ALY/REF [57]. ALY/REF also bears an N-terminal disordered arginine-rich region that resembles an RGG-box [57] and mediates both RNA binding [54, 58, 59] and the interaction with NXF1 [60]. The activation of NXF1 is proposed to be triggered by the formation of a ternary complex between ALY/REF and NXF1, in which their RG-rich disordered regions play a central role. Analogous sequences have been identified in viral proteins and also facilitate viral RNA export by bypassing canonical nuclear export pathways (Table 1).

Fragile X mental retardation protein (FMRP) is another RBP with a well-characterised, RNA-binding RGG-box (Fig. 1). Involved in translation repression in the brain [61], loss of FMRP activity leads to changes in synaptic connectivity [62], mental retardation [63–65], and may also promote onset of neurodegenerative diseases [66]. In addition to its RGG-box, FMRP contains two KH domains that contribute to RNA binding. The RGG-box of FMRP

has been shown to interact with high affinity with G-quadruplex RNA structures [67–77]. The RGG-box is unstructured in its unbound state [70, 78], but folds upon binding to a guanine-rich, structured G-quadruplex in target RNA [78] (Fig. 2). Both arginines and glycines play a key role in the function of the RGG-box and replacement of these amino acids impair RNA binding [78]. The arginine residues used to interact with RNA vary depending on the target RNA [70, 76, 78]. The FMRP RGG-box targets its own mRNA at an G-quadruplex structure that encodes the RGG-box [69]. This binding regulates alternative splicing of FMRP mRNA proximal to the G-quartet, suggesting it may auto-regulate the balance of FMRP isoforms [74]. Surprisingly, a recent transcriptome-wide study of polysome-associated FMRP found no enrichment for predicted G-quadruplex structures in the 842 high-confidence target mRNAs [79]. Another study identified FMRP binding sites enriched in specific sequence motifs, where the KH2 domains emerged as the major specificity determinants [80]. These results suggest that the role of RGG-box in this RBP could be limited to increase the overall binding affinity of the protein, supporting the sequence-specific interactions mediated by the KH2 domains. However, we cannot exclude the possibility of differential UV crosslinking efficiency of the KH2 domains and the RGG-box, which could result in biased binding signatures in CLIP studies.

A number of other RBPs use an RGG-repeat region to target G-rich and structured RNA targets and are implicated in neurological disease as well as cancer (Table 1). These RG-rich regions can mediate both unselective and specific interactions with RNA and can be involved in varied RNA metabolic processes.

Catching the RNA with a basic arm

Basic residues often cluster in RBPs to form basic patches that can contribute to RNA-binding. Analysis of mammalian RNA-binding proteomes showed that such motifs are abundant among unorthodox RBPs [25, 27]. Basic patches are normally composed of 4–8 lysines (K) or, less frequently, arginines (R), forming a highly positive and exposed interface with potential to mediate molecular interactions [25]. Basic patches can occur at multiple positions within an RBP forming islands that often flank globular domains. This suggests functional cooperation between natively structured and unstructured regions [25]. Many RBPs contain alternating basic and acidic tracts that form highly repetitive patterns with unknown function [25]. Since acidic regions are not thought to interact with RNA [58], they may be involved in other intra- or intermolecular interactions, or contribute to accessibility and compaction of the region [81].

Arginine rich motifs (ARMs) (Table 1) are probably best characterised in viral proteins. These motifs tend to

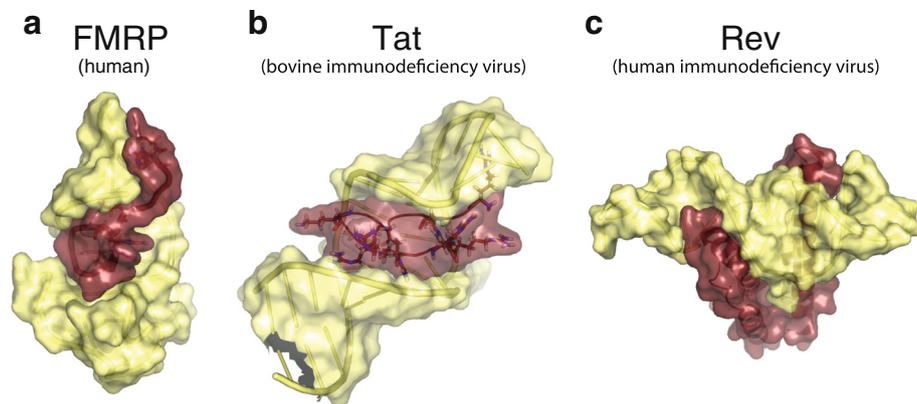


Fig. 2 Structural examples RNA-bound disordered regions. **a** The RGG-peptide of the human FMRP bound to a *in vitro*-selected guanine-rich sc1 RNA determined by NMR (PDB 2LA5) [78] **b** Basic patch of disordered bovine immunodeficiency virus (BIV) Tat forms a β -turn when interacting with its target RNA, TAR. Structure determined by NMR (PDB 1MN8) [91] **c** Dimer of the basic patch containing Rev protein of the human immunodeficiency virus (HIV) in complex with target RNA, RRE, determined by crystallography [102] (PDB 4PMI). Red, peptide; yellow, RNA. Illustrations were created using PyMol

be disordered, and when bound to RNA, range from completely disordered to ordered but flexible. Although simple in terms of amino acid composition, ARMs seem to be able to target RNAs quite diversely and often specifically [82]. Lentiviral Tat proteins (Trans-Activator of Transcription) are key regulator of viral biological cycle by promoting viral gene expression upon binding to an RNA structure present at the 5' end of the nascent viral RNA (called the trans-activation response element, TAR) [83]. Human immunodeficiency virus (HIV) Tat ARM is intrinsically disordered in its free-state [84–87]. Only one key arginine, flanked by basic amino acids, is required for specific interaction with TAR [88, 89]. Differences in the flanking basic amino acids contribute to selectivity between TARs from different viruses [90]. ARMs can accommodate different binding conformations depending on their target RNA. For example, bovine immunodeficiency virus (BIV) Tat ARM forms a beta-turn conformation upon binding to TAR [91] (Fig. 2c). Jembrana disease virus (JDV) Tat ARM can bind both HIV and BIV TARs, as well as its own TAR, but does so adopting different conformations and using different amino acids for recognition [92]. The RNA-binding disordered region of HIV Tat also mediates protein-protein interactions required for nuclear localisation [93]. Structural flexibility required to engage in diverse simultaneous or sequential RNA and protein interactions might explain why the native ARM-RNA interactions do not display very high affinity [92].

Similar to Tat proteins, lentiviral Rev auxiliary protein binds a structured RNA element (the Rev response element, RRE) present in partially-spliced and unspliced viral RNAs to facilitate nuclear export of viral RNA [94, 95]. The HIV Rev ARM was experimentally shown to be intrinsically disordered when unbound in physiological conditions [96–98] (Table 1, Fig. 1).

Disorder-to-structure transition correlates with RNA binding and the RRE-bound Rev folds into an alpha-helical structure that maintains some structural flexibility [96–100]. Rev oligomerises and binds the multiple stems of the RRE using diverse arginine contacts, which results in a high-affinity ribonucleoprotein that allows efficient nuclear export of unspliced HIV RNAs [101–103]. Interestingly, Rev can also bind in an extended conformation to *in vitro* selected RNA aptamers [104], highlighting the role of RNA secondary and tertiary structure in the conformation that Rev adopts. The RRE can also be recognised by several different *in vitro* selected R-rich peptides that include additional serine, glycine, and glutamic acid residues [105–107] — these peptides are predicted to be disordered (Table 1). A simple, single nucleotide base changes in the RRE can direct affinity towards a particular ARM [108]. These features highlight the structural malleability of the Rev ARM, and suggest that some structural flexibility is relevant for *in vivo* binding.

The basic amino acid lysine can form disordered polylysine peptides that interact with RNA. 47 proteins identified in the human RNA-binding proteome have a long poly-K patch but lack known RBDs, suggesting these motifs are good candidates for RNA binding [25]. The K-rich C-terminal tail of protein SDA1 homolog (SDAD1) is composed of 45 amino acids, including 15 K, one R, two glutamines (Q) and two asparagines (N) (Table 1, Fig. 1). It binds RNA *in vivo* with similar efficiency as a canonical domain such as RRM [58]. The human non-canonical poly(A) polymerase PAPD5, that is involved in oligoadenylation aberrant rRNAs to target them for degradation [109, 110], also lacks canonical RBDs, but its C-terminal basic patch is directly involved in binding RNA (Fig. 1, Table 1). Removal or mutation of this sequence results in impaired RNA binding and reduced catalytic activity [109].

Basic tails in RBPs share physicochemical similarities with analogous sequences in DNA-binding proteins (DBPs) [111]. In DNA-binding context, basic patches are known to endow faster association with DNA due to increased ‘capture radius’ as well as to promote hopping and sliding movements along DNA molecules [112–118]. DNA binding through basic tails seems to be sequence-independent [119] and structural studies have shown that basic residues are projected into the minor groove of the double stranded DNA helix, establishing numerous electrostatic interactions with the phosphate-sugar backbone [116, 120]. Basic patches in RBPs may modulate RNA searching and binding avidity in a similar manner.

One open question is whether basic tails can distinguish between DNA and RNA. The AT-hook, defined as G-R-P core flanked by basic arginine and/or lysine residues, binds DNA and is found in many nuclear, DNA-binding proteins [121, 122]. However, this motif has been recently shown also to bind RNA [123–126]. Furthermore, an extended AT-hook (Table 1), occurring in tens of mouse and human proteins, binds RNA with higher affinity than DNA [127]. This motif from Prostate Tumor Overexpressed 1 (PTOV1) was shown to bind structured RNA, in agreement with the previously known property of basic tails to bind in the minor groove of double stranded DNA [116, 120]. Therefore, different types of disordered sequences may be able to recognise both RNA and DNA, albeit they may have preference for one.

A role for disordered regions of RBPs in retaining RNA in membraneless granules

RNA processing and storage is often undertaken in the context of dynamic, membraneless organelles that vary in size, composition, and function. These organelles include the nucleolus, PML bodies, nuclear speckles and cajal bodies in the nucleus as well as P-bodies, stress and germ granules in the cytoplasm [128–130]. RNA granule formation relies on a spatiotemporally controlled transition from disperse “soluble” RNA and protein state to a condensed phase [131, 132]. The lack of a membrane allows a direct, dynamic and reversible exchange of components between the cytoplasm and the granule [131]. The rate of exchange and localisation of a protein within a granule can be markedly different depending on granule composition and the intrinsic properties of the protein [133–136]. RNA granules have roles in RNA localisation, stability, and translation, and perturbations in their homeostasis are hallmarks of numerous neurological disorders [137, 138].

Several recent studies have shown that disordered, low complexity regions in a number RBP have a capacity to form such granules [131, 139–141]. Different low

complexity regions can promote RNA granule formation. For example, the disordered RG-rich sequence of LAF-1 (DDX3) was demonstrated to be both necessary and sufficient to promote P-granule formation in *C. elegans* [142]. Similarly, the RG/GR and FG/GF disordered tail of human RNA helicase DDX4 (aka Vasa) aggregates *in vivo* and *in vitro* [130]. Furthermore, the [G/S]Y[G/S] and poly glutamine (polyQ) motifs, which are present in a broad spectrum of RBPs, are necessary and sufficient to cause aggregation *in vitro* and *in vivo* [139, 140, 143–146]. It remains unclear how RNA binding by these sequences influences granule formation. Illustrating this idea, the RG-rich region of LAF-1 displays direct RNA-binding activity in addition to granule formation capacity. While RNA is not required for LAF-1 driven aggregation, it increases the internal dynamics of these LAF-1 droplets, making them more fluid [142]. In yeast, formation of P-body-like granules by the Lsm4 disordered region requires the presence of RNA [147]. Notably, the biophysical properties of RBP droplets can be altered by the presence of different RNA species [148]. A recent work reports an additional layer of complexity in the interplay between nucleic acids and granules. While single-stranded DNA is retained in DDX4-induced granules, double-stranded DNA is excluded, suggesting some degree of nucleic acid selectivity [130]. Given the biophysical similarities between DNA and RNA, it is possible that granules formed by analogous low complexity sequences also retain single stranded over double stranded RNA.

Interestingly, different types of low complexity sequences may help to form different types of aggregates and ways to embed RNA. A recent study showed that while low complexity sequences promote formation of both P-bodies and stress granules in yeast, these granules differ in their dynamic properties, P-bodies displaying more dynamic/fluid phase transition than more solid-like stress granules [147]. Granule structure, composition, and age can affect the biophysical properties of the granules [135, 136]. There is considerable overlap in the composition of different RNA granules [149]. Different proportions of such components may lead to the existence of a continuum of granule types with increasingly distinct physicochemical properties. In summary, it is clear that protein disorder has a role in formation of RNA granules. The importance of direct interaction between disordered regions and RNA in the context of granules remains to be determined.

Modulating interactions between disordered regions and RNA

Post-translational modifications can modulate protein’s interaction properties [150]. A number of disordered RNA-

binding regions are known to be post-translationally modified (Table 1, Additional file 1: Figure S1) and some of these modifications can modulate RNA-binding affinity or cause local structural changes. For example, methylation of arginines of the RNA-binding RGG-box in the RNA export adapter ALY/REF reduces its affinity for RNA [151]. Arginine methylation of the RGG-box of the translational regulator FMRP affects interaction with target RNA as well as its polyribosome association [76, 152]. Also the RNA-binding basic patch of HIV protein Rev is methylated, which changes its interaction dynamics with its target RNA [153, 154]. Serine phosphorylation at the RNA-binding RS repeats of SRSF1 and DDX23 has been shown to induce a partial structuring of this region, which may impact their RNA-binding properties [36]. Assembly of RNA granules can also be modified by phosphorylation or methylation of the low complexity region [130, 155, 156]. In summary, occurrence of post-translational modifications at

disordered regions represents an additional layer of regulation of RNA binding and metabolism (Fig. 3).

In other contexts, it is known that alternative splicing can alter the sequence and function of proteins. Several global analyses have reported that short, regulatory sequences such as sites for post-translational modifications and protein-protein interactions are often subjected to alternative splicing [157–159]. Could protein-RNA interactions be regulated in a similar manner? A number alternative isoforms catalogued in large-scale studies affect RNA-binding disordered regions (Table 1, Additional file 1: Figure S1). As an illustrative example, alternative splicing of mouse ALY/REF selectively includes or excludes the RNA-binding RG-rich region, resulting in changes in its targeting to nuclear speckles and an increased cytoplasmic distribution [57, 60]. Alternative splicing affecting a region adjacent to the FMRP RGG-box influences the protein's RNA-binding activity [160], reduces its ability to associate with

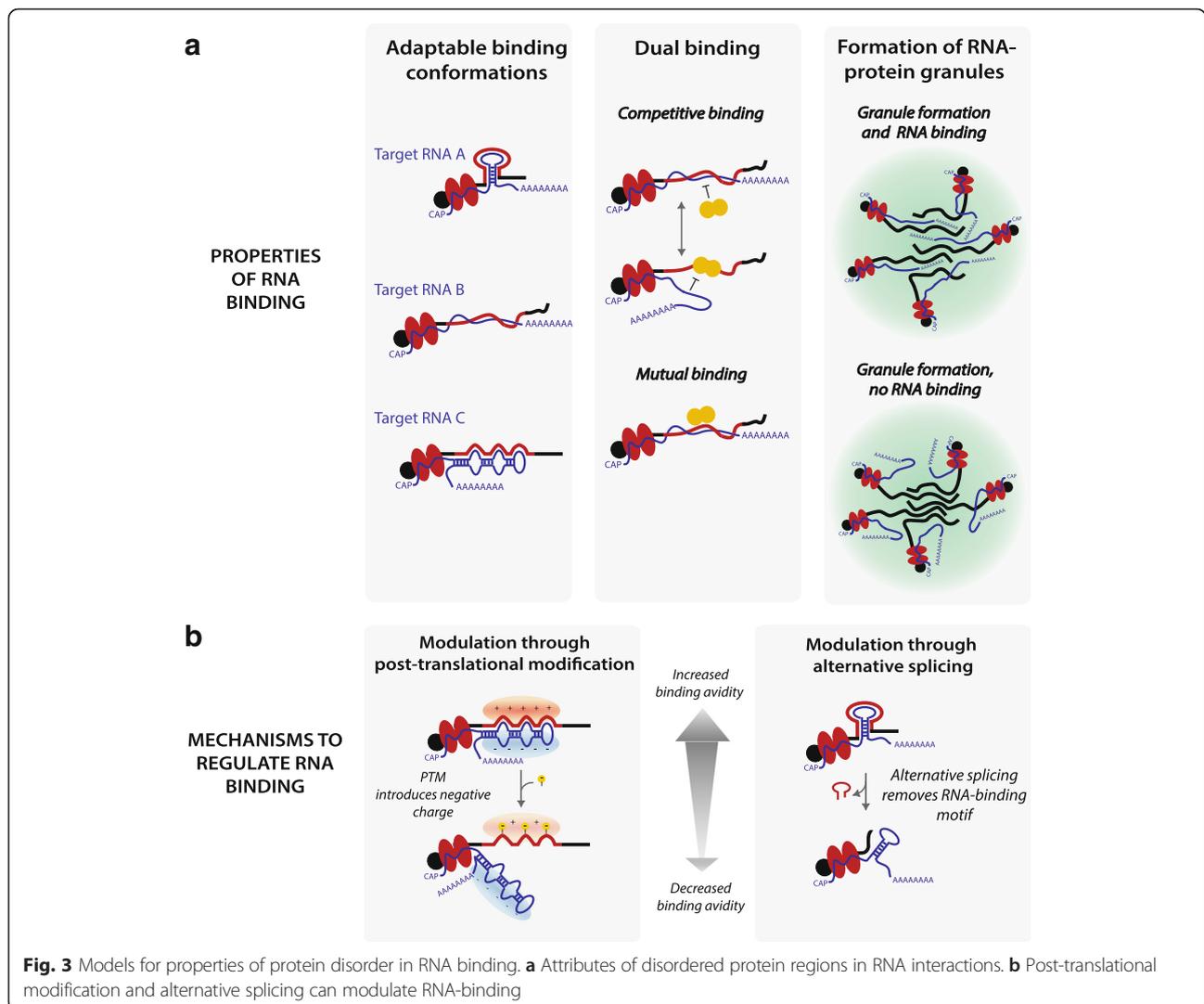


Fig. 3 Models for properties of protein disorder in RNA binding. **a** Attributes of disordered protein regions in RNA interactions. **b** Post-translational modification and alternative splicing can modulate RNA-binding

polyribosomes [161], and can also impact RGG-box methylation [162]. Another splice isoform results in ablation of the RGG-box as a result of a translational frameshift, which induces nuclear distribution of the protein [163]. Also RNA granule formation can be differentially regulated in different tissues through selective splicing isoforms including or excluding granule-forming low complexity regions [164]. Although to our knowledge a genome-wide analysis is still outstanding, these anecdotal examples hint that alternative splicing may operate to alter disorder-RNA interactions in a global manner (Fig. 3).

RNA-binding activity can also be modulated by competitive or cooperative interactions (Table 1, Fig. 3). The ability of some disordered regions to mediate protein-protein or protein-DNA interactions in addition to protein-RNA interactions could provide additional means to regulate RBP function. Therefore, disordered regions, although neglected for decades, have the potential to emerge as dynamic mediators of RNA biology.

Conclusions

Why disorder?

We have discussed the contribution of RS-, RG-, and K/R-rich, disordered regions to RNA interactions, and given examples of how they participate in co- and post-transcriptional regulation of RNA metabolism; how defects in these interactions can lead to disease; and how disorder in RBPs can be utilised by viruses during their infection cycle. Disordered regions are emerging as malleable, often multifunctional RNA-binding modules whose interactions with RNA range from non-specific to highly selective with defined target sequence or structural requirements (Fig. 3). How specificity is generated for RNA sequences or structures by disordered RNA-binding regions remains to be determined. Specific interactions with defined RNA structures have been demonstrated in some instances. It seems likely that specificity and affinity can be increased by oligomerisation and through the combinatorial modular architecture of RBPs. Disorder may be a spatially cost-effective way of encoding general affinity for RNA and/or structural flexibility to enable co-folding in presence of the target RNA, thus allowing multiple binding solutions not easily achievable by structured domains. Because disorder-mediated interaction with RNA typically relies on physicochemical properties of short stretches of sequence, they can be easily regulated through post-translational modifications. Disorder may also endow special properties such as propensity to form RNA granules and interact with other RBPs. Here we have grouped the RNA-binding disordered regions based on their amino acid composition. It is possible that other functional RNA-binding motifs with unobvious sequence patterns remain to be discovered.

Outstanding questions

Much remains to be learnt about disorder-mediated protein-RNA interactions. How do disordered regions interact with RNA? How many functionally relevant disorder-RNA interactions exist? Can more refined motifs be identified among the different classes of RNA-binding disordered regions? Are there further subclasses of motifs within RS-, RG-, basic, and other RNA-binding disordered regions with distinct binding characteristics? How is RNA binding regulated post-translationally, by alternative splicing, or by competitive interactions with other biomolecules? How do mutations in disordered regions involved in RNA binding cause disease? Fundamental principles of disorder-RNA interactions are likely to have close parallels to what has been elucidated for protein-protein and protein-DNA interactions, where disorder-mediated regulation has received much more attention over the past decade [111, 165–170]. Thus, the conceptual framework to start answering questions on the role of protein disorder in RNA binding already has a firm foundation.

Concluding statement

Structure-to-function paradigm [171] has persisted long in the field of protein-RNA interactions. In this review, we have highlighted the important role that disordered regions play in RNA binding and regulation. Indeed, the recent studies on mammalian RNA-binding proteomes place disordered regions at the centre of the still expanding universe of RNA-protein interactions. It is thus time to embark on a more systematic quest of discovery for the elusive functions of disordered protein regions in RNA biology.

Additional file

Additional file 1: Figure S1. Properties of RNA-binding, disordered proteins. Disorder and charge profiles for proteins listed in Table 1. The disordered, RNA-binding regions (RBR) are marked in blue in the left panel, and their sequence given in the right panel. Amino acid sequence, GO terms, and annotations for protein domains, isoforms, and post-translational modifications (PTMs) were extracted from UniProt [250]. Disorder was calculated using IUPred [172] using default values. Score above 0.4 indicates the region is intrinsically disordered (in physiological conditions). Charge was calculated using EMBOSS charge [251] using default values. PTMs: A, acetylation; M, methylation; P, phosphorylation; O, other. See Table 1 for literature references for each protein. (PDF 904 kb)

Abbreviations

ARM: arginine-rich motif; dsRBD: double-stranded RNA-binding domain; GAR repeat: glycine-arginine-rich repeat; KH domain: K-homology domain; RBD: RNA-binding domain; RBP: RNA-binding protein; RGG-box: arginine-glycine-glycine-box; RRM: RNA recognition motif; RS repeat: arginine-serine repeat.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AJJ and AC designed the manuscript. AIJ, MN, ID and AC wrote the manuscript. All authors have read and accepted the final version of the manuscript.

Acknowledgements

This work was supported by an MRC Career Development Award (#MR/L019434/1) granted to A.C., and by a Wellcome Trust senior research fellowship to I.D. (grant number 096144). We thank Prof. Matthias Hentze for discussion.

Received: 18 November 2015 Accepted: 21 March 2016

Published online: 06 April 2016

References

- Mitchell SF, Parker R. Principles and properties of eukaryotic mRNPs. *Mol Cell*. 2014;54:547–58.
- Müller-McNicoll M, Neugebauer KM. How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet*. 2013;14:275–87.
- Lukong KE, Chang K-W, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends Genet*. 2008;24:416–25.
- Licalosi DD, Darnell RB. Splicing regulation in neurologic disease. *Neuron*. 2006;52:93–101.
- Wurth L, Gebauer F. RNA-binding proteins, multifaceted translational regulators in cancer. *Biochim Biophys Acta*. 1849;2015:881–6.
- Neelamraju Y, Hashemikhabir S, Janga SC. The human RBPome: From genes and proteins to human disease. *J Proteomics*. 2015;127(Pt A):61–70.
- Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. *Trends Genet*. 2013;29:318–27.
- Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol*. 2007;8:479–90.
- Nicastro G, Taylor IA, Ramos A. KH-RNA interactions: back in the groove. *Curr Opin Struct Biol*. 2015;30:63–70.
- Cléry A, Allain FH-T. From structure to function of rna binding domains. In: *Madame Curie Bioscience Database* [Internet]. Austin (TX): Landes Bioscience; 2000-2013. <http://www.ncbi.nlm.nih.gov/books/NBK63528/>.
- Abbas YM, Pichlmair A, Góna MW, Superti-Furga G, Nagar B. Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature*. 2013;494:60–4.
- Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. *Nat Rev Genet*. 2014;15:829–45.
- Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume A-M, et al. The RNA-binding proteome from yeast to man harbour conserved enigmRBPs. *Nat Commun*. 2015;6:10127.
- Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res*. 2002;30:1427–64.
- Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. Diverse RNA-binding proteins interact with functionally related sets of mas, suggesting an extensive regulatory system. *PLoS Biol*. 2008;6:e255.
- Muckenthaler MU, Galy B, Hentze MW. Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. *Annu Rev Nutr*. 2008;28:197–213.
- Walden WE, Selezneva AI, Dupuy J, Volbeda A, Fontecilla Camps JC, Theil EC, et al. Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA. *Science*. 2006;314:1903–8.
- Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L, Amir-Ahmady B, Pitsch S, Black DL, Allain FH-T. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*. 2005;309:2054–7.
- Valcárcel J, Gaur RK, Singh R, Green MR. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science*. 1996;273:1706–9.
- Kiledjian M, Dreyfuss G. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J*. 1992;11:2655–64.
- Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res*. 2007;6:1882–98.
- Lobley A, Swindells MB, Orengo CA, Jones DT. Inferring function using patterns of native disorder in proteins. *PLoS Comput Biol*. 2007;3:e162.
- Tsvetanova NG, Klass DM, Salzman J, Brown PO. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS One*. 2010;5:e12671.
- Scherrer T, Mittal N, Janga SC, Gerber AP. A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. *PLoS One*. 2010;5:e15499.
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsvelde J, Hentze MW. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. 2012;149:1393–406.
- Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell*. 2012;46:674–90.
- Kwon SC, Yi H, Eichelbaum K, Föhr S, Fischer B, You KT, Castello A, Krijgsvelde J, Hentze MW, Kim VN. The RNA-binding protein repertoire of embryonic stem cells. *Nat Struct Mol Biol*. 2013;20:1122–30.
- Shepard PJ, Hertel KJ. The SR protein family. *Genome Biol*. 2009;10(10):242. doi:10.1186/gb-2009-10-10-242. Epub 2009 Oct 27. PMID: 19857271.
- Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J*. 2009;417(1):15–27. doi:10.1042/BJ20081501. PMID:19061484.
- Howard JM, Sanford JR. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip Rev RNA*. 2015;6:93–110.
- Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell*. 2009;35:1–10.
- Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol*. 2009;10:741–54.
- Hertel KJ, Graveley BR. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem Sci*. 2005;30:115–8.
- Graveley BR, Hertel KJ, Maniatis T. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J*. 1998;17:6747–56.
- Haynes C, Iakoucheva LM. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res*. 2006;34:305–12.
- Xiang S, Gapsys V, Kim H-Y, Bessonov S, Hsiao H-H, Möhlmann S, Klaukien V, Ficner R, Becker S, Urlaub H, Lührmann R, de Groot B, Zwickstetter M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure*. 2013;21:2162–74.
- Shen H, Kan JLC, Green MR. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell*. 2004;13(3):367–76. PMID:14967144.
- Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell*. 2004;16:363–73.
- Shen H, Green MR. RS domain-splicing signal interactions in splicing of U12-type and U2-type introns. *Nat Struct Mol Biol*. 2007;14:597–603.
- Jamros MA, Aubol BE, Keshwani MM, Zhang Z, Stamm S, Adams JA. Intra-domain cross-talk regulates serine-arginine protein kinase 1-dependent phosphorylation and splicing function of transformer 2β1. *J Biol Chem*. 2015;290:17269–81.
- Graveley BR. A protein interaction domain contacts RNA in the prespliceosome. *Mol Cell*. 2004;13:302–4.
- Boucher L, Ouzounis CA, Enright AJ, Blencowe BJ. A genome-wide survey of RS domain proteins. *RNA*. 2001;7:1693–701.
- Burgute BD, Peche VS, Steckelberg A-L, Glöckner G, Gaßen B, Gehring NH, Noegel AA. NKAP is a novel RS-related protein that interacts with RNA and RNA binding proteins. *Nucleic Acids Res*. 2014;42:3177–93.
- Chen D, Li Z, Yang Q, Zhang J, Zhai Z, Shu H-B. Identification of a nuclear protein that promotes NF-κappaB activation. *Biochem Biophys Res Commun*. 2003;310:720–4.
- Pajerowski AG, Nguyen C, Aghajanian H, Shapiro MJ, Shapiro VS. NKAP is a transcriptional repressor of notch signaling and is required for T cell development. *Immunity*. 2009;30:696–707.
- Chang C-K, Hsu Y-L, Chang Y-H, Chao F-A, Wu M-C, Huang Y-S, Hu C-K, Huang T-H. Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol*. 2009;83:2255–64.
- Thandapani P, O'Connor TR, Bailey TL, Richard S. Defining the RGG/RG motif. *Mol Cell*. 2013;50:613–23.

48. Kipp M, Göhring F, Ostendorp T, van Druenen CM, van Driel R, Przybylski M, Fackelmayer FO. SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol*. 2000;20:7480–9.
49. Aravind L, Koonin EV. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci*. 2000;25:112–4.
50. Romig H, Fackelmayer FO, Renz A, Ramsperger U, Richter A. Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J*. 1992;11:3431–40.
51. Xiao R, Tang P, Yang B, Huang J, Zhou Y, Shao C, Li H, Sun H, Zhang Y, Fu X-D. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell*. 2012;45:656–68.
52. Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K, Nakagawa S. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell*. 2010;19:469–76.
53. Puvvula PK, Desetty RD, Pineau P, Marchio A, Moon A, Dejean A, Bischof O. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat Commun*. 2014;5:5323.
54. Golovanov AP, Hautbergue GM, Tintaru AM, Lian L-Y, Wilson SA. The solution structure of REF2-1 reveals interdomain interactions and regions involved in binding mRNA export factors and RNA. *RNA*. 2006;12:1933–48.
55. Hautbergue GM, Hung M-L, Golovanov AP, Lian L-Y, Wilson SA. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc Natl Acad Sci U S A*. 2008;105:5154–9.
56. Braun IC, Rohrbach E, Schmitt C, Izaurralde E. TAP binds to the constitutive transport element (CTE) through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. *EMBO J*. 1999;18:1953–65.
57. Stutz F, Bachi A, Doerks T, Braun IC, Séraphin B, Wilm M, Bork P, Izaurralde E. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA*. 2000;6:638–50.
58. Strein C, Alleaume A-M, Rothbauer U, Hentze MW, Castello A. A versatile assay for RNA-binding proteins in living cells. *RNA*. 2014;20:721–31.
59. Zenklusen D, Vinciguerra P, Strahm Y, Stutz F. The yeast hnRNP-like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol*. 2001;21:4219–32.
60. Rodrigues JP, Rode M, Gatfield D, Blencowe BJ, Carmo-Fonseca M, Izaurralde E. REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc Natl Acad Sci U S A*. 2001;98:1030–5.
61. Chen E, Joseph S. Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie*. 2015;114:147–54.
62. Pfeiffer BE, Huber KM. The state of synapses in fragile X syndrome. *Neuroscientist*. 2009;15:549–67.
63. McLennan Y, Polussa J, Tassone F, Hagerman R. Fragile x syndrome. *Curr Genomics*. 2011;12:216–24.
64. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991;65:905–14.
65. De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA, Willems PJ. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet*. 1993;3:31–5.
66. Wang H. Fragile X mental retardation protein: from autism to neurodegenerative disease. *Front Cell Neurosci*. 2015;9:43.
67. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, Darnell RB, Warren ST. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*. 2001;107:477–87.
68. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*. 2001;107:489–99.
69. Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C, Moine H. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J*. 2001;20:4803–13.
70. Ramos A, Hollingworth D, Pastore A. G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*. 2003;9:1198–207.
71. Zanotti KJ, Lackey PE, Evans GL, Mihailescu M-R. Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry*. 2006;45:8319–30.
72. Menon L, Mihailescu M-R. Interactions of the G quartet forming semaphorin 3 F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res*. 2007;35:5379–92.
73. Menon L, Mader SA, Mihailescu M-R. Fragile X mental retardation protein interactions with the microtubule associated protein 1B RNA. *RNA*. 2008;14:1644–55.
74. Didiot M-C, Tian Z, Schaeffer C, Subramanian M, Mandel J-L, Moine H. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res*. 2008;36:4902–12.
75. Bole M, Menon L, Mihailescu M-R. Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA. *Mol Biosyst*. 2008;4:1212–9.
76. Blackwell E, Zhang X, Ceman S. Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA. *Hum Mol Genet*. 2010;19:1314–23.
77. Zhang Y, Gaetano CM, Williams KR, Bassell GJ, Mihailescu M-R. FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA. *RNA Biol*. 2014;11:1364–74.
78. Phan AT, Kuryavii V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D, Darnell RB, Patel DJ. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol*. 2011;18:796–804.
79. Darnell JC, Van Driesche SJ, Zhang C, Hung KYS, Mele A, Fraser CE, Stone EF, Chen C, Fak JJ, Chi SW, Licatalosi DD, Richter JD, Darnell RB. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. 2011;146:247–61.
80. Ascano M, Mukherjee N, Bandaru P, Miller JB, Nusbaum JD, Corcoran DL, et al. FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature*. 2012;492:382–6.
81. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol*. 2015;32:102–12.
82. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. *RNA*. 2005;11:1848–57.
83. Ott M, Geyer M, Zhou Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe*. 2011;10:426–35.
84. Bayer P, Kraft M, Ejchart A, Westendorp M, Frank R, Rösch P. Structural studies of HIV-1 Tat protein. *J Mol Biol*. 1995;247:529–35.
85. Calnan BJ, Biancalana S, Hudson D, Frankel AD. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev*. 1991;5:201–10.
86. Long KS, Crothers DM. Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA. *Biochemistry*. 1995;34:8885–95.
87. Aboul-ela F, Karn J, Varani G. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol*. 1995;253:313–32.
88. Calnan BJ, Tidor B, Biancalana S, Hudson D, Frankel AD. Arginine-mediated RNA recognition: the arginine fork. *Science*. 1991;252:1167–71.
89. Calabro V, Daugherty MD, Frankel AD. A single intramolecular contact mediates intramolecular stabilization of both RNA and protein. *Proc Natl Acad Sci U S A*. 2005;102:6849–54.
90. Tao J, Frankel AD. Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins. *Proc Natl Acad Sci U S A*. 1993;90:1571–5.
91. Puglisi JD, Chen L, Blanchard S, Frankel AD. Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*. 1995;270:1200–3.
92. Smith CA, Calabro V, Frankel AD. An RNA-binding chameleon. *Mol Cell*. 2000;6:1067–76.
93. Truant R, Cullen BR. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol Cell Biol*. 1999;19:1210–7.
94. Kuzembayeva M, Dilley K, Sardo L, Hu W-S. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology*. 2014;454–455:362–70.
95. Rausch JW, Le Grice SJ. HIV Rev assembly on the rev response element (RRE): a structural perspective. *Viruses*. 2015;7:3053–75.
96. Casu F, Duggan BM, Hennig M. The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. *Biophys J*. 2013;105:1004–17.
97. Tan R, Chen L, Buettner JA, Hudson D, Frankel AD. RNA recognition by an isolated alpha helix. *Cell*. 1993;73:1031–40.
98. Tan R, Frankel AD. Costabilization of peptide and RNA structure in an HIV Rev peptide-RRE complex. *Biochemistry*. 1994;33:14579–85.

99. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*. 1996;273:1547–51.
100. Wilkinson TA, Zhu L, Hu W, Chen Y. Retention of conformational flexibility in HIV-1 Rev-RNA complexes. *Biochemistry*. 2004;43:16153–60.
101. Daugherty MD, D'Orso I, Frankel AD. A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA. *Mol Cell*. 2008;31:824–34.
102. Jayaraman B, Crosby DC, Homer C, Ribeiro I, Mavor D, Frankel AD. RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev-Rev response element complex. *Elife*. 2014;3:e04120.
103. Bai Y, Tambe A, Zhou K, Doudna JA. RNA-guided assembly of Rev-RRE nuclear export complexes. *Elife*. 2014;3:e03656.
104. Ye X, Gorin A, Frederick R, Hu W, Majumdar A, Xu W, McLendon G, Ellington A, Patel DJ. RNA architecture dictates the conformations of a bound peptide. *Chem Biol*. 1999;6:657–69.
105. Harada K, Martin SS, Frankel AD. Selection of RNA-binding peptides in vivo. *Nature*. 1996;380:175–9.
106. Tan R, Frankel AD. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc Natl Acad Sci U S A*. 1998;95:4247–52.
107. Harada K, Martin SS, Tan R, Frankel AD. Molding a peptide into an RNA site by in vivo peptide evolution. *Proc Natl Acad Sci U S A*. 1997;94:11887–92.
108. Iwazaki T, Li X, Harada K. Evolvability of the mode of peptide binding by an RNA. *RNA*. 2005;11:1364–73.
109. Rammelt C, Bilen B, Zavolan M, Keller W. PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA*. 2011;17:1737–46.
110. Shcherbik N, Wang M, Lapiak YR, Srivastava L, Pestov DG. Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. *EMBO Rep*. 2010;11:106–11.
111. Vuzman D, Levy Y. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosyst*. 2012;8:47–57.
112. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A*. 2000;97:8868–73.
113. Trizac E, Levy Y, Wolynes PG. Capillarity theory for the fly-casting mechanism. *Proc Natl Acad Sci U S A*. 2010;107:2746–50.
114. Halford SE, Szczelkun MD. How to get from A to B: strategies for analysing protein motion on DNA. *Eur Biophys J*. 2002;31:257–67.
115. Elf J, Li GW, Xie XS. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*. 2007;316:1191–4.
116. Iwahara J, Zweckstetter M, Clore GM. NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc Natl Acad Sci U S A*. 2006;103:15062–7.
117. Gowers DM, Wilson GG, Halford SE. Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc Natl Acad Sci U S A*. 2005;102:15883–8.
118. Vuzman D, Levy Y. DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A*. 2010;107:21004–9.
119. Crane-Robinson C, Dragan AI, Privalov PL. The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem Sci*. 2006;31:547–52.
120. Privalov PL, Dragan AI, Crane-Robinson C, Breslauer KJ, Remeta DP, Minetti CASA. What drives proteins into the major or minor grooves of DNA? *J Mol Biol*. 2007;365:1–9.
121. Huth JR, Bewley CA, Nissen MS, Evans JN, Reeves R, Gronenborn AM, Clore GM. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Biol*. 1997;4:657–65.
122. Aravind L, Landsman D. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res*. 1998;26:4413–21.
123. Manabe T, Ohe K, Katayama T, Matsuzaki S, Yanagita T, Okuda H, Bando Y, Imaizumi K, Reeves R, Tohyama M, Mayeda A. HMGA1a: sequence-specific RNA-binding factor causing sporadic Alzheimer's disease-linked exon skipping of presenilin-2 pre-mRNA. *Genes Cells*. 2007;12:1179–91.
124. Eilebrecht S, Brysbaert G, Wegert T, Urlaub H, Benecke B-J, Benecke A. 75k small nuclear RNA directly affects HMGA1 function in transcription regulation. *Nucleic Acids Res*. 2011;39:2057–72.
125. Eilebrecht S, Wilhelm E, Benecke B-J, Bell B, Benecke AG. HMGA1 directly interacts with TAR to modulate basal and Tat-dependent HIV transcription. *RNA Biol*. 2013;10:436–44.
126. Benecke AG, Eilebrecht S. RNA-Mediated Regulation of HMGA1 Function. *Biogeosciences*. 2015;5:943–57.
127. Filarsky M, Zillner K, Araya I, Villar-Garea A, Merkl R, Längst G, Németh A. The extended AT-hook is a novel RNA binding motif. *RNA Biol*. 2015;12:864–76.
128. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc Natl Acad Sci U S A*. 2011;108:4334–9.
129. Hyman AA, Brangwynne CP. Beyond stereospecificity: liquids and mesoscale organization of cytoplasm. *Dev Cell*. 2011;21:14–6.
130. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowitz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell*. 2015;57:936–47.
131. Weber SC, Brangwynne CP. Getting RNA and protein in phase. *Cell*. 2012;149:1188–91.
132. Hyman AA, Weber CA, Jülicher F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol*. 2014;30:39–58.
133. Erickson SL, Lykke-Andersen J. Cytoplasmic mRNP granules at a glance. *J Cell Sci*. 2011;124:293–7.
134. Weil TT, Parton RM, Herpers B, Soetaert J, Veenendaal T, Xanthakis D, Dobbie IM, Halstead JM, Hayashi R, Rabouille C, Davis I. Drosophila patterning is established by differential association of mRNAs with P bodies. *Nat Cell Biol*. 2012;14:1305–13.
135. Jain S, Wheeler JR, Walters RW, Agrawal A, Barsic A, Parker R. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell*. 2016;164:487–98.
136. Lin Y, Protter DSW, Rosen MK, Parker R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol Cell*. 2015;60:208–19.
137. Buchan JR. mRNP granules. Assembly, function, and connections with disease. *RNA Biol*. 2014;11:1019–30.
138. Ramaswami M, Taylor JP, Parker R. Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell*. 2013;154:727–36.
139. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012;149:753–67.
140. Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell*. 2012;149:768–79.
141. Uversky VN, Kuznetsova IM, Turoverov KK, Zaslavsky B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Lett*. 2015;589:15–22.
142. Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC-H, Eckmann CR, Myong S, Brangwynne CP. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A*. 2015;112:7189–94.
143. Lee C, Zhang H, Baker AE, Occhipinti P, Borsuk ME, Gladfelter AS. Protein aggregation behavior regulates cyclin transcript localization and cell-cycle control. *Dev Cell*. 2013;25:572–84.
144. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res*. 2012;1462:61–80.
145. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell*. 2009;137:146–58.
146. Reijns MAM, Alexander RD, Spiller MP, Beggs JD. A role for Q/N-rich aggregation-prone regions in P-body localization. *J Cell Sci*. 2008;121:2463–72.
147. Kroschwald S, Maharana S, Mateju D, Malinowska L, Nüsse E, Poser I, Richter D, Alberti S. Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *Elife*. 2015;4:e06807.
148. Zhang H, Elbaum-Garfinkle S, Langdon EM, Taylor N, Occhipinti P, Bridges AA, et al. RNA Controls polyq protein phase transitions. *Mol Cell*. 2015;60:220–30.
149. Buchan JR, Parker R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell*. 2009;36:932–41.
150. Van Roey K, Gibson TJ, Davey NE. Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol*. 2012;22:378–85.
151. Hung M-L, Hautbergue GM, Snijders APL, Dickman MJ, Wilson SA. Arginine methylation of REF/ALY promotes efficient handover of mRNA to TAP/NXF1. *Nucleic Acids Res*. 2010;38:3351–61.

152. Stetler A, Winograd C, Sayegh J, Cheever A, Patton E, Zhang X, Clarke S, Ceman S. Identification and characterization of the methyl arginines in the fragile X mental retardation protein Fmrp. *Hum Mol Genet.* 2006; 15:87–96.
153. Hyun S, Jeong S, Yu J. Effects of asymmetric arginine dimethylation on RNA-binding peptides. *ChemBiochem.* 2008;9:2790–2.
154. Invernizzi CF, Xie B, Richard S, Wainberg MA. PRMT6 diminishes HIV-1 Rev binding to and export of viral RNA. *Retrovirology.* 2006;3:93.
155. Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, Kim J, Yun J, Xie Y, McKnight SL. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science.* 2014;345:1139–45.
156. Wang JT, Smith J, Chen B-C, Schmidt H, Rasoloson D, Paix A, Lambrus BG, Calidas D, Betzig E, Seydoux G. Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in *C. elegans*. *Elife.* 2014;3:e04591.
157. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A.* 2006;103:8390–5.
158. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A, et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell.* 2012;46:871–83.
159. Weatheritt RJ, Davey NE, Gibson TJ. Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res.* 2012;40:7123–31.
160. Denman RB, Sung YJ. Species-specific and isoform-specific RNA binding of human and mouse fragile X mental retardation proteins. *Biochem Biophys Res Commun.* 2002;292:1063–9.
161. Blackwell E, Ceman S. A new regulatory function of the region proximal to the RGG box in the fragile X mental retardation protein. *J Cell Sci.* 2011;124:3060–5.
162. Dolzhanskaya N, Merz G, Denman RB. Alternative splicing modulates protein arginine methyltransferase-dependent methylation of fragile X syndrome mental retardation protein. *Biochemistry.* 2006;45:10385–93.
163. Sittler A, Devys D, Weber C, Mandel JL. Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum Mol Genet.* 1996;5:95–102.
164. Shiina N, Nakayama K. RNA granule assembly and disassembly modulated by nuclear factor associated with double-stranded RNA 2 and nuclear factor 45. *J Biol Chem.* 2014;289:21163–80.
165. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci.* 2002;27:527–33.
166. Uversky VN, Davé V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, et al. Pathological unfolding of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev.* 2014;114:6844–79.
167. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, Davey NE. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev.* 2014; 114:6733–78.
168. Tompa P, Davey NE, Gibson TJ, Babu MM. A million peptide motifs for the molecular biologist. *Mol Cell.* 2014;55:161–9.
169. Fuxreiter M, Simon I, Bondos S. Dynamic protein-DNA recognition: beyond what can be seen. *Trends Biochem Sci.* 2011;36:415–23.
170. Dyson HJ. Roles of intrinsic disorder in protein-nucleic acid interactions. *Mol Biosyst.* 2012;8:97–104.
171. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* 1999;293:321–31.
172. Dosztányi Z, Csiszák V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol.* 2005;347:827–39.
173. Xu X, Yang D, Ding J-H, Wang W, Chu P-H, Dalton ND, Wang H-Y, Birmingham JR, Ye Z, Liu F, Rosenfeld MG, Manley JL, Ross J, Chen J, Xiao R-P, Cheng H, Fu X-D. ASF/SF2-regulated CaMKII δ alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell.* 2005;120:59–72.
174. Ge H, Zuo P, Manley JL. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell.* 1991;66:373–82.
175. Nayler O, Strätling W, Bourquin JP, Stagljar I, Lindemann L, Jasper H, et al. SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res.* 1998;26:3542–9.
176. David CJ, Boyne AR, Millhouse SR, Manley JL. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev.* 2011;25:972–83.
177. Chang C-K, Sue S-C, Yu T-H, Hsieh C-M, Tsai C-K, Chiang Y-C, Lee S-J, Hsiao H-H, Wu W-J, Chang W-L, Lin C-H, Huang T-H. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci.* 2006;13:59–72.
178. Tunnicliffe RB, Hautbergue GM, Wilson SA, Kalra P, Golovanov AP. Competitive and cooperative interactions mediate RNA transfer from herpesvirus saimiri ORF57 to the mammalian export adaptor ALYREF. *PLoS Pathog.* 2014;10:e1003907.
179. Thandapani P, Song J, Gandin V, Cai Y, Rouleau SG, Garant J-M, Boisvert F-M, Yu Z, Perreault J-P, Topisirovic I, Richard S. Aven recognition of RNA G-quadruplexes regulates translation of the mixed lineage leukemia protooncogenes. *Elife.* 2015;4. doi:10.7554/eLife.06234. PMID:26267306.
180. Ina S, Tsunekawa N, Nakamura A, Noce T. Expression of the mouse Aven gene during spermatogenesis, analyzed by subtraction screening using Mvh-knockout mice. *Gene Expr Patterns.* 2003;3:635–8.
181. Solomon S, Xu Y, Wang B, David MD, Schubert P, Kennedy D, Schrader JW. Distinct structural features of caprin-1 mediate its interaction with G3BP-1 and its induction of phosphorylation of eukaryotic translation initiation factor 2 α , entry to cytoplasmic stress granules, and selective interaction with a subset of mRNAs. *Mol Cell Biol.* 2007;27:2324–42.
182. Shiina N, Shinkura K, Tokunaga M. A novel RNA-binding protein in neuronal RNA granules: regulatory machinery for local translation. *J Neurosci.* 2005;25:4420–34.
183. Takahama K, Kino K, Arai S, Kurokawa R, Oyoshi T. Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein. *FEBS J.* 2011;278:988–98.
184. Shaw DJ, Morse R, Todd AG, Eggleton P, Lorson CL, Young PJ. Identification of a self-association domain in the Ewing's sarcoma protein: a novel function for arginine-glycine-glycine rich motifs? *J Biochem.* 2010;147:885–93.
185. Belyanskaya LL, Gehring PM, Gehring H. Exposure on cell surface and extensive arginine methylation of ewing sarcoma (EWS) protein. *J Biol Chem.* 2001;276:18681–7.
186. Belyanskaya LL, Delattre O, Gehring H. Expression and subcellular localization of Ewing sarcoma (EWS) protein is affected by the methylation process. *Exp Cell Res.* 2003;288:374–81.
187. Araya N, Hiraga H, Kako K, Arao Y, Kato S, Fukamizu A. Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochem Biophys Res Commun.* 2005;329:653–60.
188. Vasilyev N, Polonskaia A, Darnell JC, Darnell RB, Patel DJ, Serganov A. Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc Natl Acad Sci U S A.* 2015;112: E5391–400.
189. Menon RP, Gibson TJ, Pastore A. The C terminus of fragile X mental retardation protein interacts with the multi-domain Ran-binding protein in the microtubule-organising centre. *J Mol Biol.* 2004;343:43–53.
190. Masuda A, Takeda J-I, Okuno T, Okawara B, Ito M, Ishigaki S, Sobue G, Ohno K. Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev.* 2015;29:1045–57.
191. Nakaya T, Alexiou P, Maragkakis M, Chang A, Mourelatos Z. FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA.* 2013;19:498–509.
192. Takahama K, Oyoshi T. Specific binding of modified RGG domain in TLS/FUS to G-quadruplex RNA: tyrosines in RGG domain recognize 2'-OH of the riboses of loops in G-quadruplex. *J Am Chem Soc.* 2013; 135:18016–9.
193. Rappsilber J, Friesen WJ, Paushkin S, Dreyfuss G, Mann M. Detection of arginine dimethylated peptides by parallel precursor ion scanning mass spectrometry in positive ion mode. *Anal Chem.* 2003;75:3107–14.
194. Mears WE, Rice SA. The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation. *J Virol.* 1996;70:7445–53.
195. Souki SK, Gershon PD, Sandri-Goldin RM. Arginine methylation of the ICP27 RGG box regulates ICP27 export and is required for efficient herpes simplex virus 1 replication. *J Virol.* 2009;83:5309–20.
196. Corbin-Lickfett KA, Souki SK, Cocco MJ, Sandri-Goldin RM. Three arginine residues within the RGG box are crucial for ICP27 binding to herpes simplex virus 1 GC-rich sequences and for efficient viral RNA export. *J Virol.* 2010;84:6367–76.
197. Viphakone N, Cumberbatch MG, Livingstone MJ, Heath PR, Dickman MJ, Catto JW, Wilson SA. Luszp4 defines a new mRNA export pathway in cancer cells. *Nucleic Acids Res.* 2015;43:2353–66.

198. Viphakone N, Hautbergue GM, Walsh M, Chang C-T, Holland A, Folco EG, Reed R, Wilson SA. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun.* 2012;3:1006.
199. Ghisolfi L, Joseph G, Amalric F, Erard M. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J Biol Chem.* 1992;267:2955–9.
200. Bouvet P, Diaz JJ, Kindbeiter K, Madjar JJ, Amalric F. Nucleolin interacts with several ribosomal proteins through its RGG domain. *J Biol Chem.* 1998;273:19025–9.
201. Kanhoush R, Beenders B, Perrin C, Moreau J, Bellini M, Penrad-Mobayed M. Novel domains in the hnRNP G/RBMX protein with distinct roles in RNA binding and targeting nascent transcripts. *Nucleus.* 2010;1:109–22.
202. Dichmann DS, Fletcher RB, Harland RM. Expression cloning in *Xenopus* identifies RNA-binding proteins as regulators of embryogenesis and Rbmx as necessary for neural and muscle development. *Dev Dyn.* 2008;237:1755–66.
203. Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol.* 2012;14:318–28.
204. Tsend-Ayush E, O'Sullivan LA, Grütznér FS, Onnebo SMN, Lewis RS, Delbridge ML, Marshall Graves JA, Ward AC. RBMX gene is essential for brain development in zebrafish. *Dev Dyn.* 2005;234:682–8.
205. Shashi V, Xie P, Schoch K, Goldstein DB, Howard TD, Berry MN, Schwartz CE, Cronin K, Sliwa S, Allen A, Need AC. The RBMX gene as a candidate for the Shashi X-linked intellectual disability syndrome. *Clin Genet.* 2015;88:386–90.
206. Mazeyrat S, Saut N, Mattei MG, Mitchell MJ. RBMY evolved on the Y chromosome from a ubiquitously transcribed X-Y identical gene. *Nat Genet.* 1999;22:224–6.
207. Schliephake AW, Rethwilm A. Nuclear localization of foamy virus Gag precursor protein. *J Virol.* 1994;68:4946–54.
208. Yu SF, Edelmann K, Strong RK, Moebes A, Rethwilm A, Linnal ML. The carboxyl terminus of the human foamy virus Gag protein contains separable nucleic acid binding and nuclear transport domains. *J Virol.* 1996;70:8255–62.
209. Hamann MV, Mullers E, Reh J, Stanke N, Effantin G, Weissenhorn W, Lindemann D. The cooperative function of arginine residues in the Prototype Foamy Virus Gag C-terminus mediates viral and cellular RNA encapsidation. *Retrovirology.* 2014;11:87.
210. Mullers E, Uhlig T, Stirnagel K, Fiebig U, Zentgraf H, Lindemann D. Novel Functions of Prototype Foamy Virus Gag Glycine- Arginine-Rich Boxes in Reverse Transcription and Particle Morphogenesis. *J Virol.* 2011;85:1452–63.
211. Zhang P, Abdelmohsen K, Liu Y, Tominaga-Yamanaka K, Yoon J-H, Ioannis G, Martindale JL, Zhang Y, Becker KG, Yang IH, Gorospe M, Mattson MP. Novel RNA- and FMRP-binding protein TRF2-S regulates axonal mRNA transport and presynaptic plasticity. *Nat Commun.* 2015;6:8888.
212. Mitchell TRH, Glenfield K, Jeyanthan K, Zhu X-D. Arginine methylation regulates telomere length and stability. *Mol Cell Biol.* 2009;29:4918–34.
213. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell.* 2009;35:403–13.
214. Zhang P, Casaday-Potts R, Precht P, Jiang H, Liu Y, Pazin MJ, Mattson MP. Nontelomeric splice variant of telomere repeat-binding factor 2 maintains neuronal traits by sequestering repressor element 1-silencing transcription factor. *Proc Natl Acad Sci U S A.* 2011;108:16434–9.
215. Lapointe CP, Wickens M. The nucleic acid-binding domain and translational repression activity of a *Xenopus* terminal uridylyl transferase. *J Biol Chem.* 2013;288:20723–33.
216. Weeks KM, Ampe C, Schultz SC, Steitz TA, Crothers DM. Fragments of the HIV-1 Tat protein specifically bind TAR RNA. *Science.* 1990;249:1281–5.
217. Frankel AD. Fitting peptides into the RNA world. *Curr Opin Struct Biol.* 2000;10:332–40.
218. Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science.* 1992;257:76–80.
219. Kumar S, Maiti S. Effect of different arginine methylations on the thermodynamics of Tat peptide binding to HIV-1 TAR RNA. *Biochimie.* 2013;95:1422–31.
220. Xie B, Invernizzi CF, Richard S, Wainberg MA. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat Interactions with both cyclin T1 and the Tat transactivation region. *J Virol.* 2007;81:4226–34.
221. Deng L, La Fuente De C, Fu P, Wang L, Donnelly R, Wade JD, Lambert P, Li H, Lee CG, Kashanchi F. Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones. *Virology.* 2000;277:278–95.
222. D'Orso I, Frankel AD. Tat acetylation modulates assembly of a viral-host RNA-protein transcription complex. *Proc Natl Acad Sci U S A.* 2009;106:3101–6.
223. Kaehlcke K, Dorr A, Hetzer Egger C, Kiermer V, Henklein P, Schnoelzer M, Loret E, Cole PA, Verdin E, Ott M. Acetylation of Tat defines a cyclinT1-independent step in HIV transactivation. *Mol Cell.* 2003;12:167–76.
224. Sherpa C, Rausch JW, Le Grice SFJ, Hammarskjöld M-L, Rekosh D. The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Res.* 2015;43:4676–86.
225. Ansel-McKinney P, Scott SW, Swanson M, Ge X, Gehrke L. A plant viral coat protein RNA binding consensus sequence contains a crucial arginine. *EMBO J.* 1996;15:5077–84.
226. Kan JH, Andree PJ, Kouijzer LC, Mellema JE. Proton-magnetic-resonance studies on the coat protein of alfalfa mosaic virus. *Eur J Biochem.* 1982;126:29–33.
227. Anosova I, Melnik S, Tripsianes K, Kateb F, Grummt I, Sattler M. A novel RNA binding surface of the TAM domain of TIP5/BAZ2A mediates epigenetic regulation of rRNA genes. *Nucleic Acids Res.* 2015;43:5208–20.
228. Gu L, Frommel SC, Oakes CC, Simon R, Grupp K, Gerig CY, Bär D, Robinson MD, Baer C, Weiss M, Gu Z, Schapira M, Kuner R, Sülthmann H, Provenzano M, Yaspo M-L, Brors B, Korbel J, Schlomm T, Sauter G, Eils R, Plass C, Santoro R. BAZ2A (TIP5) is involved in epigenetic alterations in prostate cancer and its overexpression predicts disease recurrence. *Nat Genet.* 2014;47:22–30.
229. Zhang J, Lieu YK, Ali AM, Penson A, Reggio KS, Rabadan R, et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci U S A.* 2015;112:E4726–34.
230. Daubner GM, Cléry A, Jayne S, Stevenin J, Allain FH-T. A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J.* 2012;31:162–74.
231. Tsuda K, Someya T, Kuwasako K, Takahashi M, He F, Unzai S, Inoue M, Harada T, Watanabe S, Terada T, Kobayashi N, Shirouzu M, Kigawa T, Tanaka A, Sugano S, Güntert P, Yokoyama S, Muto Y. Structural basis for the dual RNA-recognition modes of human Tra2- β RRM. *Nucleic Acids Res.* 2011;39:1538–53.
232. Cléry A, Jayne S, Benderska N, Dominguez C, Stamm S, Allain FH-T. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2- β 1. *Nat Struct Mol Biol.* 2011;18:443–50.
233. Beil B, Screation G, Stamm S. Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. *DNA Cell Biol.* 1997;16:679–90.
234. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, Mittag T, Taylor JP. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* 2015;163:123–33.
235. Nadler SG, Merrill BM, Roberts WJ, Keating KM, Lisbin MJ, Barnett SF, Wilson SH, Williams KR. Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry.* 1991;30:2968–76.
236. Mayeda A, Munroe SH, Cáceres JF, Krainer AR. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* 1994;13:5483–95.
237. Weighardt F, Biamonti G, Riva S. Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J Cell Sci.* 1995;108(Pt 2):545–55.
238. Preitner N, Quan J, Nowakowski DW, Hancock ML, Shi J, Tcherkezian J, Young-Pearse TL, Flanagan JG. APC is an RNA-binding protein, and its interactome provides a link to neural development and microtubule assembly. *Cell.* 2014;158:368–82.
239. Saldaña-Meyer R, González-Buendía E, Guerrero G, Narenda V, Bonasio R, Recillas-Targa F, Reinberg D. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* 2014;28:723–34.
240. Szöllösi E, Bokor M, Bodor A, Perczel A, Klement E, Medzihradsky KF, Tompa K, Tompa P. Intrinsic structural disorder of DF31, a drosophilaprotein of chromatin decondensation and remodeling activities. *J Proteome Res.* 2008;7:2291–9.

241. Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA, Reinberg D. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* 2010;24:2615–20.
242. Zhang Y, Yang X, Gui B, Xie G, Zhang D, Shang Y, Liang J. Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J Biol Chem.* 2011;286:42414–25.
243. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat rna to the mouse X chromosome. *Science.* 2008;322:750–6.
244. Yue MM, Lv K, Meredith SC, Martindale JL, Gorospe M, Schuger L. Novel RNA-binding protein P311 binds eukaryotic translation initiation factor 3 subunit b (eIF3b) to promote translation of transforming growth factor β 1-3 (TGF- β 1-3). *J Biol Chem.* 2014;289:33971–83.
245. Fernandez-Chamorro J, Pineiro D, Gordon JMB, Ramajo J, Francisco-Velilla R, Macias MJ, et al. Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res.* 2014;42:5742–54.
246. Dimaano C. RNA Association defines a functionally conserved domain in the nuclear pore protein Nup153. *J Biol Chem.* 2001;276:45349–57.
247. Ball JR. The RNA binding domain within the nucleoporin Nup153 associates preferentially with single-stranded RNA. *RNA.* 2004;10:19–27.
248. Bonasio R, Lecona E, Narendra V, Voigt P, Parisi F, Kluger Y, Reinberg D. Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. *Elife.* 2014;3:e02637.
249. Zoabi M, Nadar-Ponniiah PT, Khoury-Haddad H, Usaj M, Budowski-Tal I, Haran T, Henn A, Mandel-Gutfreund Y, Ayoub N. RNA-dependent chromatin localization of KDM4D lysine demethylase promotes H3K9me3 demethylation. *Nucleic Acids Res.* 2014;42:13026–38.
250. UniProt Consortium: UniProt: a hub for protein information. *Nucleic Acids Res* 2015, 43(Database issue):D204–12.
251. Rice P, Longden I, Bleasby A: EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000, 16:276–277.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

