Received: 18 May 2020 | Revised: 30 September 2020 | Accepted: 1 October 2020

DOI: 10.1002/bies.202000118

BioEssays WILEY

THINK AGAIN

Insights & Perspectives

RNA-protein interactions: Central players in coordination of regulatory networks

Alexandros Armaos[1,2] | **Elsa Zacco**[2] | **Natalia Sanchez de Groot**[1] | **Gian Gaetano Tartaglia**[1,2,3,4]

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain

[2] Center for Human Technologies, Istituto Italiano di Tecnologia, Genova, Italy

[3] Department of Biology 'Charles Darwin', Sapienza University of Rome, Rome, Italy

[4] Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Correspondence
Natalia Sanchez de Groot, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain.
Email: natalia.sanchez@crg.es
Gian Gaetano Tartaglia, Centre for Human Technologies, Italian Institute of Technology (IIT), Enrico Melen 83, 16152 Genova and Sapienza University, Viale Aldo Moro, 00185, Roma, Italy
Email: gian.tartaglia@iit.it

Funding information
European Research Council, Grant/Award Numbers: RIBOMYLOME_309545, ASTRA_855923, H2020 projects IASIS_727658, INFORE_825080; Spanish Ministry of Economy and Competitiveness, Grant/Award Number: BFU2017-86970-P; European Union's Horizon 2020

Abstract

Changes in the abundance of protein and RNA molecules can impair the formation of complexes in the cell leading to toxicity and death. Here we exploit the information contained in protein, RNA and DNA interaction networks to provide a comprehensive view of the regulation layers controlling the concentration-dependent formation of assemblies in the cell. We present the emerging concept that RNAs can act as scaffolds to promote the formation ribonucleoprotein complexes and coordinate the post-transcriptional layer of gene regulation. We describe the structural and interaction network properties that characterize the ability of protein and RNA molecules to interact and phase separate in liquid-like compartments. Finally, we show that presence of structurally disordered regions in proteins correlate with the propensity to undergo liquid-to-solid phase transitions and cause human diseases. Also see the video abstract here https://youtu.be/kfpqibsNfS0

KEYWORDS
cell homeostasis, cell regulation, Interaction network, intrinsically disordered protein, liquid phase separation, RNA binding protein

INTRODUCTION

Healthy cellular growth and development require a tight control of gene regulation, the process determining if, when and how abundantly a certain gene is expressed. Gene regulation requires the crosstalk among protein-protein, protein-DNA and protein-RNA interaction networks to guarantee stability and functionality to all biochemical processes in the cell. Dysregulation of any of these interactions networks can impair cellular functions and lead to cell death.[1] This phenomenon is particularly critical for genes whose alteration in abundance[2] cause toxicity.[3]

The formation of macromolecular complexes is concentration-dependent and requires specific stoichiometric proportions to function correctly.[4] For this reason, macromolecular assemblies are affected[5] when the abundances of their components change without control. If not modulated by interactions with other molecules such as nucleic acids,[6] formation of large protein complexes can result either in aberrant aggregation, due to the inability to maintain solubility,[7] or in toxic gain of function, when additional partners are attracted.[8]

The study of protein-RNA interactions suggested an important regulatory role, played by transcripts, in coordinating the formation of protein assemblies.[7,9] Here, we discuss the formation and

composition of interaction networks established at the DNA (transcriptional layer of regulation), RNA (translational layer of regulation) and protein (post-translational layer of regulation) levels. With this review, we wish to show that these findings can be integrated in the framework of regulatory networks controlling biological functions.

PROTEIN-DNA, PROTEIN-RNA AND PROTEIN-PROTEIN INTERACTION COORDINATE REGULATORY NETWORKS

With the aim of understanding how proteins are regulated from the interaction network point of view,[10] we analyzed the properties of four different genesets that, due to their cellular roles and physicochemical properties, establish a large number of contacts with other macromolecules(Figure 1A; Table 1).

Transcription factors (TFs) form large macromolecular assemblies

Recent chromatin immunoprecipitation and high-throughput sequencing experiments have provided genome-wide details of transcription factors binding sites, revealing important information on TFs activities in human cells.[11] TFs form large complexes with their protein partners[12] acting in a combinatorial way to regulate common target genes through specific contacts with DNA elements.[13] A subgroup of TFs has also RNA-binding ability and examples include Mothers against decapentaplegic homolog SMAD[14] and Lamin B Receptor[15] (Figure 1 B and Figure S1).

RNA-binding proteins (RBPs) bind in a combinatorial way

Similar to TFs, RBPs build large complexes with other proteins to regulate the stability and translation of transcripts,[16] as well as processes related to RNA processing such as splicing and polyadenylation.[17,18] The number of RBP partners is particularly large and in some cases, such as for instance the ribosome, different arrangements of the constitutive components, or *combinatorial*,[19] results in high heterogeneity and specialization of the whole translation system (Figure 1B and Figure S1).

Heat shock proteins (HSPs) interact with a large part of the proteome

Highly conserved in evolution and abundant in the cell, HSPs interact with a large number of proteins and are key elements in the post-translational layer.[20] The heat shock protein 70 (Hsp70) has an essential role of "molecular chaperone" assisting in protein folding, disaggregation, and degradation.[21] Hsp70 is a physical platform for the binding of client proteins, other chaperones and co-chaperones,[22] and is central in protein-protein interaction networks.[23] Intriguingly, Hsp70 interacts with its own mRNA[24] and has a disordered C-terminal region of ~10 kDa[25] (highly conserved across species and containing the Glu-Glu-Val-Asp regulatory motif), which indicates that some physicochemical properties are shared with the intrinsically disordered protein (IDP) and RBP class (Figure 1B and Figure S1).

Intrinsically disordered proteins (IDPs) are widespread

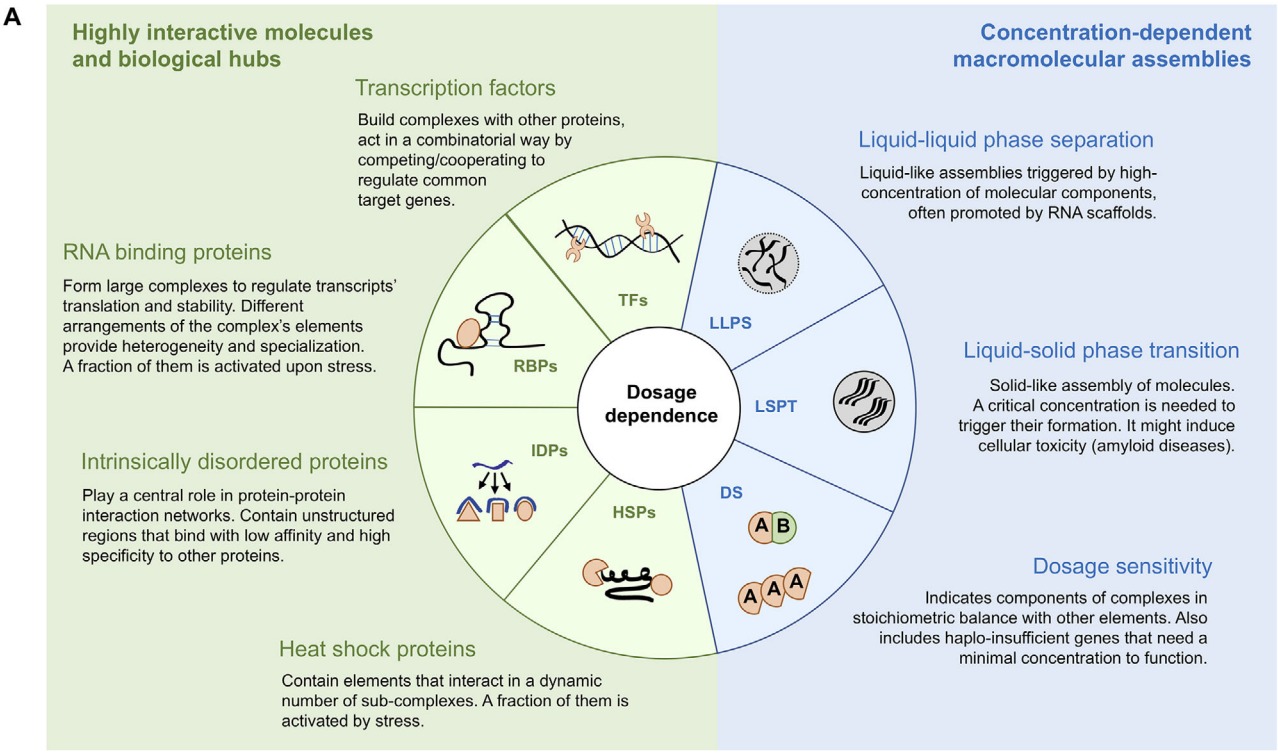
Genes coding for IDPs play a central role in protein-protein interaction networks. IDPs have the property of containing unstructured regions that bind with low affinity and high specificity to other proteins[10] forming multiple and transient complexes.[26] Classic examples of IDPs are the 40S and 60S components of the ribosome[27] and several neuronal proteins such as SNCA,[28] that binds to multiple proteins involved in synaptic vesicle formation and dopamine level control[29] (Figure 1B and Figure S1). Interestingly, a fraction of IDPs include RBPs and TFs (Figure 1B and Figure S1). Among them we can find splicing factors such as the Transformer-2 protein homolog alpha TRA2A[30] and the transcriptional repressor Ying and Yang (YY1) that can bind to both DNA and RNA.[31]

INTERACTION NETWORKS ARE ORCHESTRATED BY MASTER REGULATORS

We collected the main features of TFs, RBPs, HSPs and IDPs in the context of interaction networks. While IDPs and HSPs are almost exquisitely involved in the post-translational layer of gene regulation (protein-protein networks),TFs (protein-DNA networks) and RBPs (protein-RNA networks) act respectively at the transcriptional and translational layers, although RBPs should be considered also active at the pre-translational layer, especially for transcripts splicing, polyadenylation and localization.[16]

HSPs, RBPs and IDPs are highly abundant and associated with a large number of protein-protein interactions

With respect to the rest of the proteome (P), TFs, RBPs, HSPs and IDPs show a significantly larger number of protein interactions, in accordance with their role of master regulators (Figure 2A; Table 1; Supplementary Information). As IDPs, RBPs and HSPs are highly abundant (Figure 2B) and active in the cell at all times (Supplementary Information), changes in their concentration are expected to produce strong effects, because they can give rise to stoichiometric imbalance of protein complexes.[32] By contrast, TFs are poorly abundant and increase their expression only in specific phases of cell development or under external *stimuli*.[33,34] Yet, if uncontrolled,



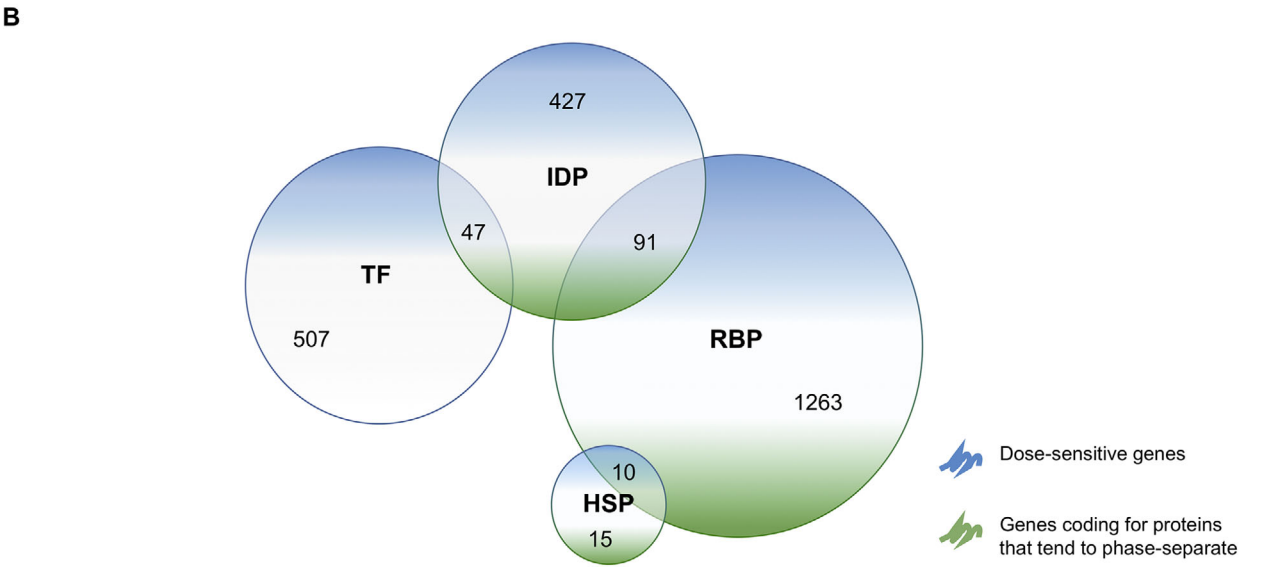
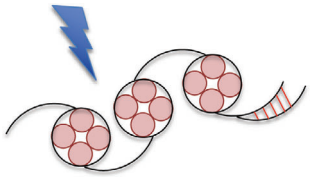
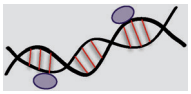
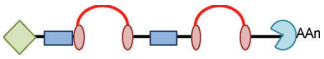
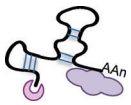
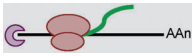


FIGURE 1 Dosage-dependent assembly. (A) Seven protein groups whose concentration dependence is linked to specific cellular functions and interaction network properties. Green: proteins forming high number of complexes and performing as "hubs" in multiple cellular processes; blue: proteins whose concentration increases and stoichiometric unbalances can either trigger the formation of large macromolecular assemblies or prevent their formation. (B) Associations among the four classes of gene examined. Circles represent gene classes (coding for TF, IDPs, RBPs or HSPs), with a diameter proportionate to the gene set size; overlap between two circles indicates the number of genes identified in both classes. Green: portion of genes of a set coding for proteins that tend to phase-separate; blue: portion of dose-sensitive genes within each set. The figure represents only the population overlaps >1%. For a more complete information, see Supplementary Figure S1 and Table S1

changes in TFs expression can lead to severe imbalance of cell functions. For example, the abundance of RE-1 silencing transcription factor REST, a repressive transcription factor active in neurons, physiologically increases during aging[35] but in Alzheimer's disease patients the expression levels are constant, which indicates impairments of functional networks.[36] Specifically in the case of Parkinson's disease, stimulation of REST expression by trichostatin A results in increased cell fitness, as shown using in vitro (SH-SY5Y cells) and *in vivo* (nigrostriatal dopaminergic neurons) models of the disease.[37]

TABLE 1 Cellular regulatory levels and their associated interaction networks

Regulatory levels	Network and function	Data source
Signal 	Stimulus and signal transmission *Heat shock*	**Gene cards**
Transcription 	Protein-DNA interactions *Transcription factors regulation*	GTRD JASPAR
Post-transcription 	RNA secondary structure *Protection and interaction*	Parallel analysis of RNA structure PARS
	Protein-RNA interaction *RNA processing, RNA expression, Translation regulation*	ENCODE RNA interactome capture
Translation 	Protein-Protein interactions *Translation regulation*	PaxDb BioGRID Genecards AmyPro
Post-translation 	Stoichiometric requirements *Protein abundance regulation*	

Regulatory levels in the cell (left). To achieve this regulation, specific "hub" proteins have to interact in a coordinated manner through specific functional networks (middle). Information relative to the interaction networks discussed in this work are reported (right; Supporting Information)

HSPs, RBPs and IDPs are tightly regulated by protein-DNA networks

Analysis of transcriptional networks indicates that HSPs, RBPs, IDPs and TFs have comparable degrees of regulation at the transcriptional layer (Figure 2C; Supplementary Information).[38] This finding indicates that these genes are tightly controlled at all levels and interaction networks must act in great synchrony. RBPs and HSPs are the most regulated (Figure 2C), which is required to optimize the response to external *stimuli*, such as environmental changes[39] and stress.[40] DNA damages, for instance, are known to down-regulate the transcription of the anti-apoptotic RBP Staufen2, resulting in activation of cell death pathways.[41] Several combinatorial associations of TFs also regulate HSPs depending on *stimuli*.[42] TFs are activated upon stress,[43] and, for example, interferon-γ treatment increases the levels of Hsp70 through STAT-1 that interacts with HSF1, antagonizing the effects of other TFs such asSTAT-3.[44] More in general, TFs act hierarchically to exert control in developmental programs[45] and in many cases, such as for BRCA1[46] and P53,[47] they exploit auto-regulatory feed-back loops to control their expression. IDPs are present in the TF machinery and, as in the case of the Mediator co-activator complex, can induce conformational changes that allow the mediator complex and RNA polymerase II to work.[48] Importantly, transcripts encoding IDPs in humans tend to have higher proportions of predicted miRNA target sites and higher mRNA decay rates, indicative of their tight regulation.[49]

HSPs, RBPs and IDPs are tightly regulated by protein-RNA networks

Although current literature does not agree on whether RNA and protein levels are correlated in the cell,[50,51] we found that this is indeed the case for HSPs, RBPs, IDPs and TFs(Figure 2D; same results were observed for RNA levels of K562 and HepG2 cell lines; Supplementary Information) for which a strong degree of cooperation exists between pre-translational and post-translational networks.[52] We note that RBPs cause changes in gene expression that are about one order of magnitude smaller than those caused by TFs.[53] Yet, RBP scan significantly alter gene expression, as shown by several IDPs such as SNCA, that have their expression and translation tightly modulated by RBP networks.[54] Given the *spectrum* of conditions in which they operate (infection, inflammation, different toxins), HSPs have the strongest regulation at post-transcriptional level (Figure 2E;

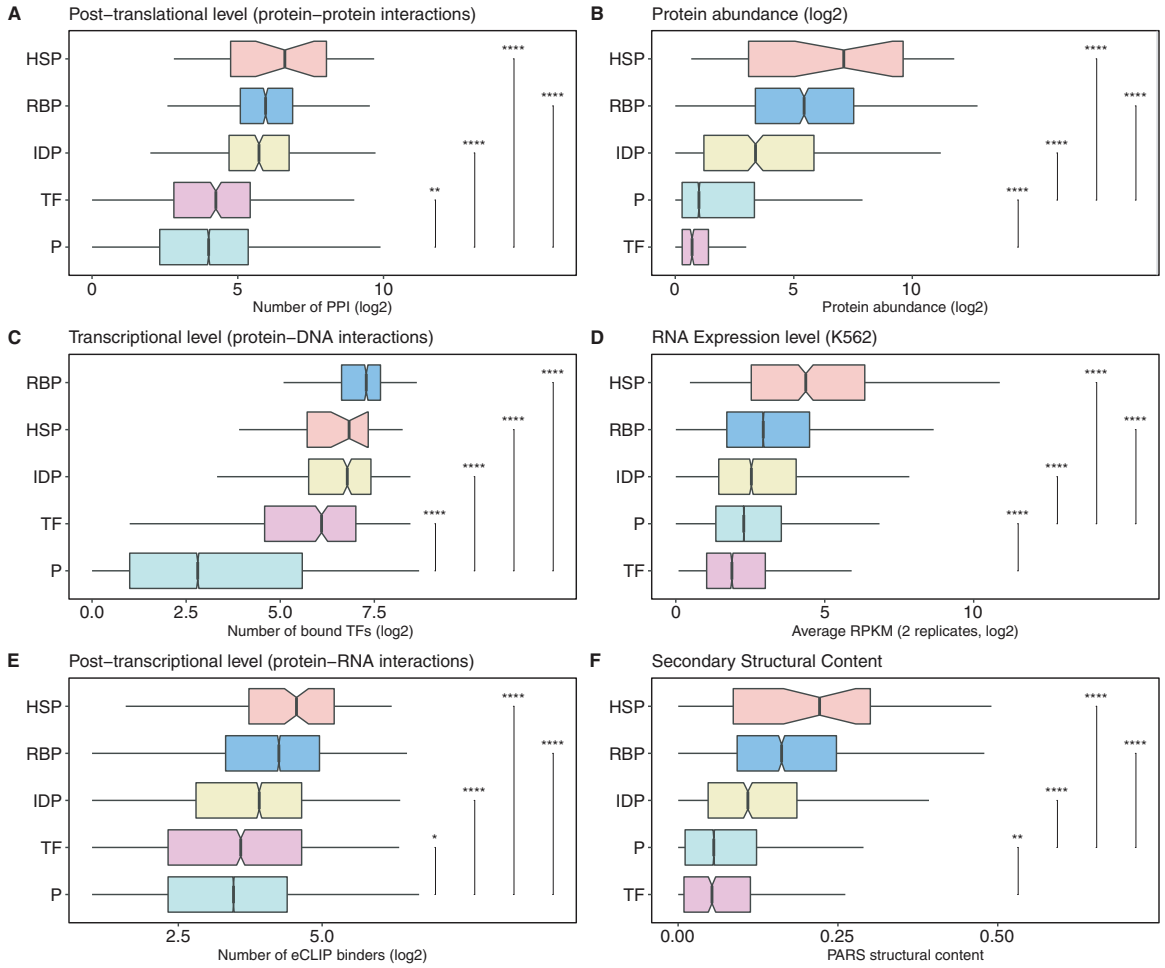


FIGURE 2 Coordination of protein-protein, protein-DNA and protein-RNA networks for master regulators. Box plots comparing properties of "hub" proteins: intrinsically disordered proteins (IDPs), heat shock proteins (HSPs), transcription factors (TFs), RNA-binding proteins (RBPs) and as a control the rest of the proteome (P) (Supporting Information). (A) Protein-protein network or post-translational layer of regulation; (B) Protein abundance; (C) Protein-DNA network or transcriptional layer of regulation; (D) RNA abundance; (E) Protein-RNA network or post- transcriptional layer of regulation; (F) RNA secondary structure content(Supporting Information). Notably, (F) correlates with (E), as reported in our recent work[6] and follows the same trend presented in (A, B and D) panels. The boxes show the interquartile range (IQR), the central line represents the median, the whiskers add 1.5 times the IQR to the 75 percentile (box upper limit) and subtract 1.5 times the IQR from the 25 percentile (box lower limit). The significance was calculated using Kolmogorov-Smirnov test (*$P< 0.05$; **$P< 0.01$; ***$P< 0.001$; ****$P< 0.0001$)

Supplementary Information). Similarly to TFs, RBPs control their expression by acting on their RNAs through auto-regulatory loops and specific RBP networks.[55,56]

STRUCTURE-DRIVEN RNA INTERACTIVITY IS AN EMERGING PROPERTY OF PROTEIN-RNA NETWORKS

The amount of RNA structure influences interaction with proteins. Gladfelter and co-workers observed that secondary structure defines the ability of certain RNAs to self-associate and phase separate with the poly-Q protein Whi3.[9] More recently, it has been reported the presence of stable structured elements flanking single stranded RNA sequences is recognized by SARS-Co-2 nucleocapsid protein to trigger formation of large protein-RNA assembly,[57,58] which influences viral packing. From a transcriptome point of view, Seemann and co-workers, observed a tight relationship between protein binding and conservation of structural elements in RNAs.[59] Some long non-coding RNAs such as *NEAT1*[60] and *XIST*[61] have been shown to exploit their structured domains to scaffold protein assemblies. It should be mentioned that there is a present and active debate about structural differences between coding and non-coding transcripts.[62]

In agreement with the examples above we recently reported that the number of RBP-RNA interactions (Figure 2E) correlates with the amount of double-stranded regions especially in coding transcripts (Figure 2F).[6] The origin of this relationship, observed with a number of different experimental approaches, is that double-stranded regions increase the amount of structure in RNAs, reducing its intrinsic flexibility . By contrast, RNAs targeting complementary regions in nucleic

acids such as antisense, microRNAs and of long intergenic non-coding RNAs (lincRNAs) display the smallest amount of structure.[6] We note that while for each amino acid residue there are two torsional degrees of freedom, RNA dimensionality is much greater– for each nucleotide residue there are seven independent torsion angles: six backbone torsional angles and one angle that describes the rotation of the base relative to the sugar.[63] Thus, an RBP can bind to a specific nucleotide region, single- or double-stranded, more tightly if the RNA partner contains a certain amount of structured regions. Presence of a folded structure in a RNA molecule favors the formation of stable and well-defined binding sites with functional roles and, in turn, evolutionary selection for RBP binding.[6,59] We stress that our observation does not suggest that protein binding sites and double-stranded regions coincide. If a specific interaction occurs in a small loop at the end of a long stem, the overall region can be considered enriched in double-stranded nucleotides, although the exact binding is in a single-stranded region.[64] Thus, structured RNA also means that loops and single stranded RNA regions between double stranded RNA stems exhibit less conformational flexibility and provide attachment sites for single stranded RNA-binding protein.

Our observation was originally based on the analysis of RNA structures measured *in vitro*,[6] which could differ from the *in vivo* one for the action of RNA-binding proteins and other molecules.[65,66] Although the mechanisms of structure formation *in vivo* are still poorly characterized,[67] previous analysis suggests a prevalence of single-stranded regions [68] and conservation of double-stranded regions in specific cases.[69] Indeed, in the complex cellular environment, RNA undergoes a number of modifications such as methylation that can influence RNA structures.[70,71] Despite the increase in single-stranded regions *in vivo*,[65,66] we found that the correlation between amount of double-stranded regions and number of protein interactions is still conserved *in vivo*, which further supports the general validity of the trend (Figure S2).

In summary, our analysis reveals that the main players of all layers of gene regulation have a large number of interactions (Figure 2). For these genes, the RNA and protein expression levels correlate with the number of protein-protein, protein-DNA and protein-RNA contacts and are proportional to the number of structured regions in the encoded transcripts, which constitute a specific signature to be further investigated.

RNA MEDIATES MACROMOLECULAR ASSEMBLY IN THE CELL

Recent breakthroughs indicate the existence of different types of biological assemblies that fall within a *spectrum* of matter states, spanning from the highly reactive liquid-like state to the nearly-inert solid amyloid fibrils.[72,73] The assemblies can involve molecules of different or equal nature (as in the case of amyloid fibrils) and RNA molecules have been identified as either inductors or inhibitors of complex formation.[74] Since structured regions in RNA molecules promote interactions with proteins,[6] transcripts with large amount of double-stranded regions have high propensity to attract proteins in large complexes.[30] These transcripts act as scaffolds for the formation of ribonucleoprotein complexes that are able to phase separate in the nucleus or cytoplasm.[75]

RNA controls liquid-liquid phase separation (LLPS)

Triggered by a high-concentration of the constitutive components and favored by RNA scaffolds,[75,76] liquid-liquid phase separation (LLPS) is a rather common process in the cell and consists in the formation of large ribonucleoprotein assemblies (Figure 1; Table 1).[77,78] A number of phase-separated organelles have been shown to contain specific mixtures of RNAs and RBPs that are difficult to characterize due to their intrinsic lability.[79] Indeed, these assemblies exchange elements with the surrounding environment and adapt to the cellular conditions in a dynamic way.[6] One of the membraneless organelles requiring phase separation to form is the ribosome.[80] Before its export to the cytoplasm, ribosomal components self-assemble in the *nucleoli*. Processing of ribosomal RNAs initiates in the dense fibrillar part of the nucleolus and continues in the granular component,[81] where the RNAs attracts ribosomal proteins and forms phase-separated subunits with them.[82] Stress granules (SGs) are another example of phase-separated organelle containing both protein and RNA molecules. SG form in the cell simultaneously to translation inhibition, contribute to regulation of gene expression in physiological conditions and are involved in pathologies such as neurodegeneration.[83]

RBP proteins undergo liquid-to-solid phase transition (LSPT)

SG proteins are particularly prone to LLPS (>25% of our LLPS database is composed of SG proteins)[30,84] and in specific conditions can undergo liquid-to-solid phase transition (LSPT),[85] which in biomedical literature is often referred to as aggregation[86] or deposition.[30] LSPT often results in formation of amyloid fibrils[87] that induce cellular toxicity when mutations[88] or chemical changes[89] alter the structure or concentration of proteins, such as in the case of the SG proteins fused in Sarcoma (FUS),[90] TDP-43[91] and SOD1.[92] LSPT is rather a wide spread phenomenon[93] tightly linked to protein abundance.[7] Indeed, biophysical experiments[85] indicate that there is a critical concentration, specific for each protein,[94,95] above which aggregation is favored and formation of amyloid fibrils promoted.[96] Moreover, many proteins have strong tendency to form amyloid structures,[97] in which they arrange themselves into fibrils composed by stranded β-sheets[93] (Supplementary Information).

LLPS and LSPT are cases of dosage sensitivity (DS)

LLPS and LSPT are intimately linked to dosage sensitivity.[98] Indeed, the broad class of dosage-sensitive (DS) proteins includes components

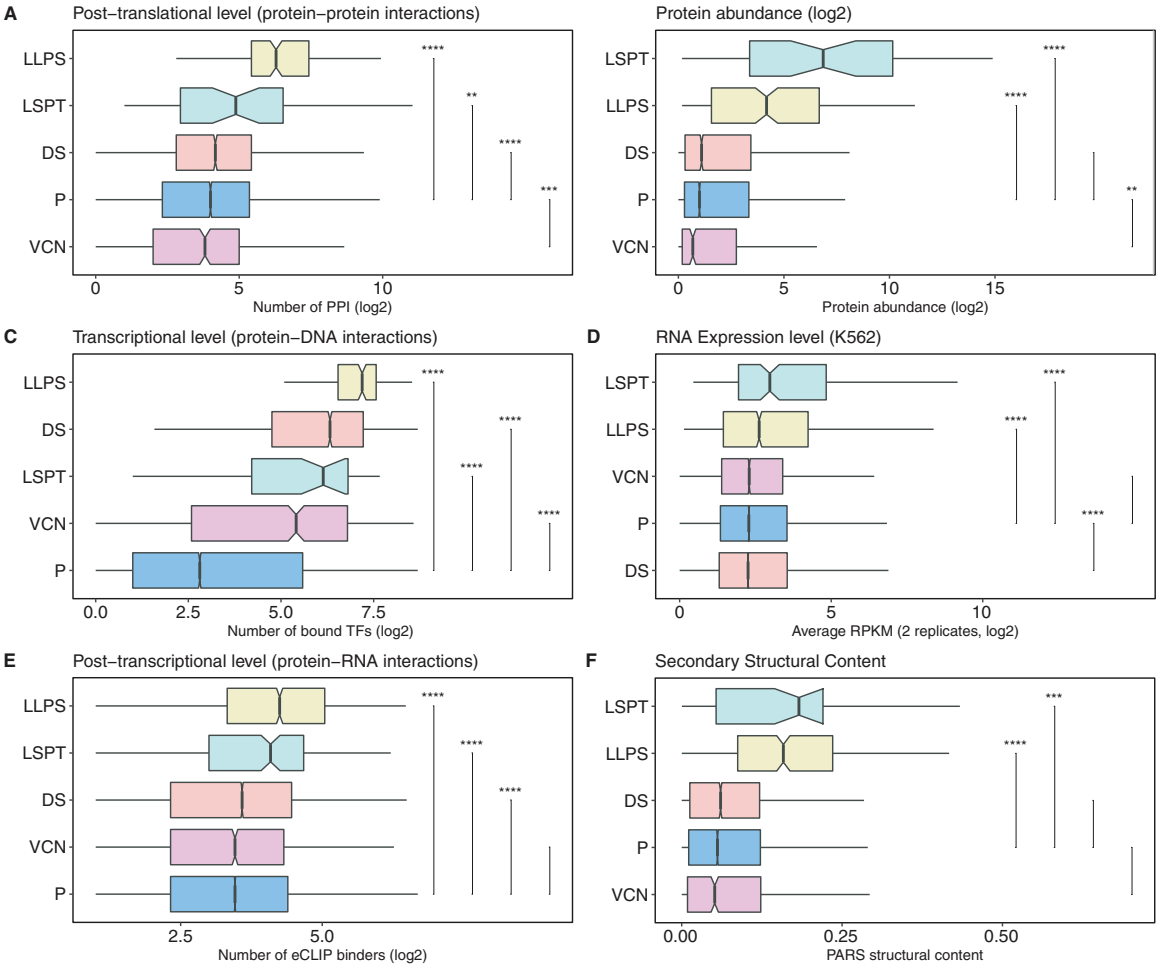


FIGURE 3 Coordination of protein-protein, protein-DNA and protein-RNA networks for the control of concentration-dependent interactions. Box plots comparing properties of 'dosage-dependent' proteins: liquid-liquid phase separation (LLPS), liquid-solid phase transition (LSPT), dosage sensitive (DS), variable copy number (VCN) proteins and the rest of the proteome (P) (Supporting Information). (A) Protein-protein network or post-translational layer of regulation; (B) Protein abundance; (C) Protein-DNA network or transcriptional layer of regulation; (D) RNA abundance; (E) Protein-RNA network or post-transcriptional layer of regulation; (F) RNA secondary structure content (Supporting Information). Notably, (F) correlates with (E), as reported in our recent work[6] and follows the same trend presented in (A, B, and C) panels. The boxes show the interquartile range (IQR), the central line represents the median, the whiskers add 1.5 times the IQR to the 75 percentile (box upper limit) and subtract 1.5 times the IQR from the 25 percentile (box lower limit). The significance was calculated using Kolmogorov-Smirnov test (*$P< 0.05$; **$P< 0.01$; ***$P< 0.001$; ****$P< 0.0001$)

of complexes[99] that are in stoichiometric balance with other elements such as the ribosome; haplo-insufficient proteins that need a minimal concentration to function, such as Fragile X Mental Retardation Protein (FMRP); proteins aggregating at high concentration, such as SNCA.[28]

PHASE SEPARATION IN THE CELL IS CONTROLLED BY SPECIFIC INTERACTIONS

We are particularly interested in the events leading to rewiring of interaction networks. Indeed, formation of complexes is a physiological event triggered by concertation and establishment of intermolecular contacts (Figure 1; Table 1). Aberrant assembly or perturbation in the composition of complexes is associated with multiple human diseases.[100,101] With the aim of understanding how dosage dependence and assembling are regulated from a network point of view, we analyzed the abundance and interactions of specific protein sets.

LLPS and LSPT proteins are highly abundant and associated with a large number of protein-protein interactions

LLPS and LSPT proteins have more protein partners than other DS proteins linked to pathogenicity[2] (Figure 3A). Indeed, SG[30,84] and amyloid proteins[101] are known to sequester numerous other proteins, interfering with the correct functioning of the cell. By contrast, proteins associated with high variable copy number (VCN) and no

toxicity establish a much smaller number of protein-protein interactions (Figure 3A), which suggests limited functionality within the cell. VCN genes have a large copy number (e.g., due to duplication) and are associated with a phenotype that is clinically interpreted as "benign".[2] Our observations suggest that one major determinant of cell toxicity is the co-aggregation of proteins, in large part IDPs, that recruit partners through structurally disordered elements (Supplementary Information).[101]

LSPT in general, and amyloid proteins in particular, are highly abundant in the cell[7] (Figure 3B) and their concentrations have been shown to often exceed the ones required for their solubility.[102] This property, called "super-saturation",[103,104] is especially related to the function of neuronal genes that are normally highly abundant in order to control a large number of synaptic processes. Examples include, for instance, alpha-synuclein, protein that mediates dopamine neurotransmission,[54] tau, a microtubule-associated protein involved in Alzheimer's disease,[105] and Huntingtin, part of the polyQ protein family.[106] Similarly, LLPS proteins are highly expressed to "sense" environmental changes to which the cell must react quickly[107] (Figure 3B). Accordingly, SGs form to protect RNAs and proteins upon external insults.[108] By contrast, DS proteins are poorly expressed (Figure 3B) and are tightly regulated by in several growth and developmental processes[108] to avoid detrimental dysfunctions such as those linked to aneuploidy disorders.[8]

LLPS and LSPT are tightly regulated by protein-DNA networks

The class of VCN genes shows low abundance levels, which, in agreement with the poor amount of protein partners (Figure 3A), suggests restricted functionality, as confirmed by the lack of transcriptional regulation compared with the DS class (Figure 3C). DS genes are under strong TF control, which is in line with their link to disease when dysregulated.[109] LLPS and, to some extent, LSPT are controlled at the transcriptional layer.[110] SGs in particular were originally described as structures into which the TF HSF1 concentrates upon heat shock.[111] Importantly, a fraction of SG proteins is represented by RBPs that, in turn, regulate the processing of many genes, including tumor suppressors and on co-proteins.[112] For instance, the stress granule-associated protein GTPase-activating protein (SH3 domain)-binding protein 2 (G3BP2) regulates breast tumor initiation through the stabilization of Squamous cell carcinoma antigen recognized by T-cells 3 (SART3) mRNA.[113]

LLPS and LSPT are tightly regulated by protein-RNA networks

Following the trend identified for TF, IDP,RBP and HSP genes (Figure 2D,F), protein and RNA levels of VCN, DS, LLPS and LSPT classes show correlation (Figure 3D,F), which suggests that the post-translational level is tightly synchronized with the pre-translational one. LLPS and LSPT are particularly regulated at the post-transcriptional level (Figure 3E), in line with the RBP property of forming "small-world" networks,[114] in which functionally-related proteins act as circuits or "regulons".[115] One minimal regulon is represented by the interaction that specific RBPs can establish with their own mRNAs to control their expression through feedback loops.[55,56] For instance, Fragile Mental Retardation protein (FMPR)[116] and Tar DNA-binding protein (TDP43),[117] both belonging to the LLPS group, bind to their cognate RNAs to limit the abundance of their protein levels and avoid aggregation.[118] Regulons containing larger RBP communities are intense object of study[17,119] and LSPT proteins such as SNCA could be involved in negative feedback loops to control their expression levels.[55,56] DS proteins (Figure 3E) in general show significant post-transcriptional regulation, which is in line with the fact that several proto-oncogenes, cytokines, cell cycle regulators and regulatory proteins involved in tumorigenesis and cancer progression are under RBP control.[120]

As previously shown in the analysis of HSP, IDP, RBPs and TF genes (Figure 2E,F), the number of RBP interactions correlate with the amount of RNA secondary structure (Figure 3E,F). While VCN genes show poor amount of double-stranded regions, which confirms their poor functional role at the protein-protein (Figure 3A), protein-DNA (Figure 3C) and protein-RNA levels (Figure 3E), DS and, in particular, LLPS and LSPT proteins have strong structural content. This observation is in agreement with previous reports indicating that LSPT-coding RNA have larger 5′ UTRs to better control their translation through RBP interactions.[110] The 3′ UTRs of genes coding for LLPS proteins are the longest of all classes analyzed, indicating tight post-transcriptional regulation.

STRUCTURAL DISORDER AND RNA-BINDING ABILITY ARE KEY DETERMINANTS OF PATHOLOGICAL PHASE TRANSITIONS

Dysregulation of biological networks and toxicity are triggered by environmental changes as well as mutations in DNA molecules producing proteins that interact less efficiently or aggregate.[7,121,122] Despite the strong coordination between the different layers of regulation (Figure 2 and Figure 3), protein-DNA, protein-RNA and protein-protein interactions are not always able to perform their functions. When their synchronization is disrupted, devastating diseases such as cancer and neurodegeneration take place.[123,124] To understand how properties of the master regulators are linked to pathology, we investigated the link between the occurrence of diseases and the HSP, IDP, RBP and TF gene content of VCN, DS, LLPS and LSPT classes. We found that VCN is the class with the lowest amount of disease-related genes and LSPT has the strongest association with pathology (Supplementary Information). In addition, the LSPT class shows a high content of disease-related genes, in agreement with its involvement in neurodegeneration and other human pathologies.[73]

Alteration in composition and spatio-temporal formation of protein macromolecular assemblies can cause cellular disturbance.[81,125]

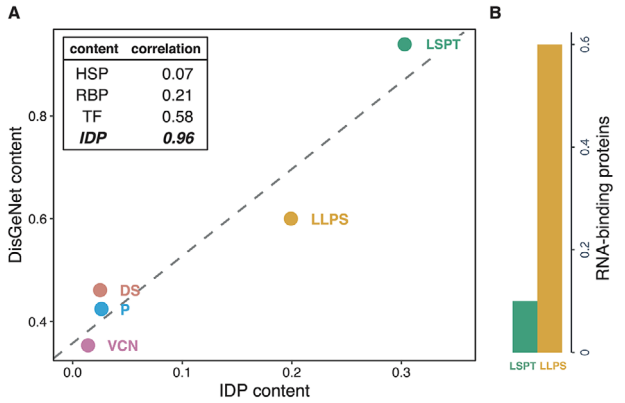


FIGURE 4 Intrinsically disordered protein (IDP) content links disease to formation of concentration-dependent assemblies. Analysis of liquid-liquid phase separation (LLPS), liquid-solid phase transition (LSPT), dosage sensitive (DS) and variable copy number (VCN) proteins. (A) Comparison beteween the fraction of anotated IDPs and the fraction of disease associated genes in the different protein sets. Inset, there is a 0.96 correlation (Pearson's) between number of IDPs and proteins reported to be associated with disease. (B)In addition to number of IDPs, another major discriminant between LLPS and LSPT is the number of RBP proteins composing the set. Pearson's correlations are reported. Disease genes from https://www.disgenet.org/

The intrinsic ability of a protein to undergo LSPT and form amyloid fibrils often results in a loss or gain of function and is linked to a higher probability to form highly reactive immature protein forms (e.g., protofibrils).[72,73] Indeed, amyloid formation propensity is related to the propensity to progress into aberrant oligomerization states.[126]

While HSP, RBP and TF contents do not show an appreciable link with the number of disease-related genes, the IDP class shows a remarkable correlation (Figure 4A). This result is particularly interesting because misexpressed, misprocessed and dysregulated IDPs are highly prone to engage in promiscuous interactions with other proteins, DNA or RNA molecules, causing pathological states.[127] Moreover, an "evolutionary biochemistry" approach based on molecular modelling and NMR experiments suggests that interactions in a proteome start with low-affinity IDPs that become structured and specific through progressive mutations.[128] Thus, the promiscuity of IDP and their ability to interact at all levels in protein-protein, protein-DNA and protein-RNA layers could be the origin of the versatility of gene networks but also its weakness.

Interestingly, the LSPT group is depleted in RNA-binding ability with respect to LSPS genes (Figure 4B). In agreement, it has been observed a decrease in liquid-like behavior after the disruption of RNA-binding domains,[129] this may favor the formation of interactions leading to solid-like aggregation.[62,86] This observation also suggests that transcripts could have the role of "solubilizers," a characteristic that may be lost in disease-related genes.[6,130] Indeed, the interaction with RNA can modulate the dynamics and material state of ribonucleoprotein complexes. This ability is associated with the fact that RNAs are flexible and multivalent favoring the interaction with multiple RBPs.

Multivalency, as the capacity to form several interactions, is a critical property associated to LSPS, whereas LSPT can be triggered by the highly aggregation-prone prion-like domain (PrLD).[131] Interestingly, a subgroup of highly interacting IDPs containing PrLD and RNA-binding domains have been found significantly enriched in phase-separating proteins.[30,84] In agreement with this observation, nucleic-acid binding abilities and disorder were reported to be enriched in SGs.[74] These properties have been recently used to build computational algorithms able to discern between LSPS and no LSPS proteins.[30,84]

CONCLUSIONS AND OUTLOOK

With the aim of studying how cells organize the assembly of complexes, we analyzed their interaction networks and regulatory layers (Figure 1).

We discussed the finding that RNA molecules rich in double stranded regions are highly prone to associate with proteins (Figure 2).[6] Our observations unveil the existence of "feedback loops" (proteins change RNA structure[65,70] and RNA induces protein assembly[30,61]) that regulate cellular events.[56,132] In addition to protein-protein and protein-RNA interactions,[108,133] the RNA-RNA network should be better investigated to achieve a full description of the post-transcriptional layer of regulation.[67]

The enrichment of RNA-binding proteins in the liquid-liquid phase separation set (e.g., stress granules) and intrinsically disordered proteins in the liquid-to-solid phase transition set (e.g., amyloids) suggests that protein assembly is affected by the presence of RNA molecules as well as unfolded polypeptide regions such as prion-like domains (Figure 3).[74] In fact, RNAs, not originally included in the "protein-only hypothesis" of prion propagation,[134,135] could be regarded as critical factors influencing protein aggregation.[74] More generally, RNAs[133] are potent organizers of the material state[30,61] and are able to alter the overall solubility of macromolecular complexes.[6,130]

Mutations in transcription factors, RNA-binding proteins, heat-shock chaperones and intrinsically disordered proteins can dramatically modify the coordination between regulatory networks.[103] The occurrence of human disease is often accompanied with a decrease of specific interactions with nucleic acids and an increase in structural disorder of proteins (Figure 4). The fact that intrinsically disordered proteins are prone to form promiscuous interactions with proteins and especially RNA and DNA should be further investigated.[124,127] Indeed, understanding the co-evolution between protein and nucleic-acids interactions will be key to unravel the mechanisms governing cell homeostasis.

ACKNOWLEDGMENTS

We thank all members of the Tartaglia and Gustincich laboratories and Dr. Marc Torrent Burgas. The research leading to these results has been supported by European Research Council (RIBOMYLOME_309545 and ASTRA_855923), the H2020 projects IASIS_727658 and INFORE_825080, the Spanish Ministry of Economy and Competitiveness BFU2017-86970-P, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754490 within the MINDED project.We

also acknowledge support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya.

CONFLICT OF INTEREST

The authors do not have any conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Gian Gaetano Tartaglia https://orcid.org/0000-0001-7524-6310

REFERENCES

1. Pascal, R., Pross, A., & Sutherland, J. D. (2013). Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol*, *3*, 130156.
2. Rice, A. M., & McLysaght, A. (2017). Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat. Commun.*, *8*, 1–11.
3. Birchler, J. A., & Veitia, R. A. (2012). Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. USA*, *109*, 14746–14753.
4. Moriya, H. (2015). Quantitative nature of overexpression experiments. *Mol. Biol. Cell*, *26*, 3932–3939.
5. Protter, D. S. W., & Parker, R. (2016). Principles and properties of stress granules. *Trends Cell Biol.*, *26*, 668–679.
6. Sanchez de Groot, N., Armaos, A., Graña-Montes, R., Alriquet, M., Calloni, G., Vabulas, R. M., & Tartaglia, G. G. (2019). RNA structure drives interaction with proteins. *Nat. Commun.*, *10*, 3246.
7. Tartaglia, G. G., Pechmann, S., Dobson, C. M., & Vendruscolo, M. (2007). Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.*, *32*, 204–206.
8. Oromendia, A. B., & Amon, A. (2014). Aneuploidy: Implications for protein homeostasis and disease. *Dis. Models Mech.*, *7*, 15–20.
9. Langdon, E. M., Qiu, Y., Niaki, A. G., McLaughlin, G. A., Weidmann, C. A., Gerbich, T. M., … Gladfelter, A. S. (2018). mRNA structure determines specificity of a polyQ-driven phase separation. *Science*, *360*, 922–927.
10. MacNeil, L. T., & Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.*, *21*, 645–657.
11. Jiang, S., Mortazavi, A. Integrating ChIP-seq with other functional genomics data.
12. Morgunova, E., & Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.*, *47*, 1–8.
13. Cirillo, D., Botta-Orfila, T., & Tartaglia, G. G. (2015). By the company they keep: Interaction networks define the binding ability of transcription factors. *Nucl. Acids Res.*,*43*, e125–e125, gkv607.
14. Dickey, T. H., & Pyle, A. M. (2017). The SMAD3 transcription factor binds complex RNA structures with high affinity. *Nucleic Acids Res.*, *45*, 11980–11988.
15. Cirillo, D., Blanco, M., Armaos, A., Buness, A., Avner, P., Guttman, M., … Tartaglia, G. G. (2017). Quantitative predictions of protein interactions with long noncoding RNAs: To the Editor. *Nat. Methods*, *14*, 5–6.
16. Marchese, D., de Groot, N. S., Lorenzo Gotor, N., Livi, C. M., & Tartaglia, G. G. (2016). Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip Rev RNA*, *7*, 793–810.
17. Zanzoni, A., Spinelli, L., Ribeiro, D. M., Tartaglia, G. G., & Brun, C. (2019). Post-transcriptional regulatory patterns revealed by protein-RNA interactions. *Sci. Rep.*, *9*
18. Kishore, S., Luber, S., & Zavolan, M. (2010). Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Briefings Funct. Genomics*, *9*, 391–404.
19. Genuth, N. R., & Barna, M. (2018). Heterogeneity and specialized functions of translation machinery: From genes to organisms. *Nat. Rev. Genet.*, *19*, 431–452.
20. Feder, M. E., & Hofmann, G. E. (1999) Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and Ecological Physiology. *Annu. Rev. Physiol.*, *61*: 243–282.
21. Tartaglia, G. G., Dobson, C. M., Hartl, F. U., & Vendruscolo, M. (2010). Physicochemical determinants of chaperone requirements. *J. Mol. Biol.*, *400*, 579–588.
22. Stetler, R. A., Gan, Y., Zhang, W., Liou, A. K., Gao, Y., Cao, G., & Chen, J. (2010). Heat shock proteins: Cellular and molecular mechanisms in the central nervous system. *Prog. Neurobiol.*, *92*, 184–211.
23. Freilich, R., Arhar, T., Abrams, J. L., & Gestwicki, J. E. (2018). Protein-Protein Interactions in the Molecular Chaperone Network. *Acc. Chem. Res.*, *51*, 940–949.
24. Balakrishnan, K., & De Maio, A. (2006). Heat shock protein 70 binds its own messenger ribonucleic acid as part of a gene expression self-limiting mechanism. *Cell Stress Chaperones*, *11*, 44–50.
25. Smock, R. G., Blackburn, M. E., & Gierasch, L. M. (2011). Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity. *J. Biol. Chem.*, *286*, 31821–31829.
26. Babu, M. M., van der Lee, R., de Groot, N. S., & Gsponer, J. (2011). Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Biol.*, *21*, 432–440.
27. Peng, Z., Oldfield, C. J., Xue, B., Mizianty, M. J., Dunker, A. K., Kurgan, L., & Uversky, V. N. (2014). A creature with a hundred waggly tails: Intrinsically disordered proteins in the ribosome. *Cell. Mol. Life Sci.*, *71*, 1477–1504.
28. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F., & Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.*, *380*, 425–436.
29. Burré, J. (2015). The synaptic function of α-synuclein. *J. Parkinsons Dis.*, *5*, 699–713.
30. Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Ponti, R. D., Severijnen, L.-A., … Tartaglia, G. G. (2018). An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.*, *25*, 3422–3434.e7.e7.
31. Agostini, F., Cirillo, D., Bolognesi, B., & Tartaglia, G. G. (2013). X-inactivation: Quantitative predictions of protein interactions in the Xist network. *Nucleic Acids Res.*, *41*, e31.
32. Birchler, J. A., Riddle, N. C., Auger, D. L., & Veitia, R. A. (2005). Dosage balance in gene regulation: Biological implications. *Trends Genet.*, *21*, 219–226.
33. Iwafuchi-Doi, M., & Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development*, *143*, 1833–1837.
34. Molina, N., Suter, D. M., Cannavo, R., Zoller, B., Gotic, I., & Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc. Natl. Acad. Sci. USA*, *110*, 20563–20568.
35. Garcia-Manteiga, J. M., D'alessandro, R., & Meldolesi, J. (2020). News about the role of the transcription factor REST in neurons: From physiology to pathology. *Int. J. Mol. Sci.*, 21
36. Baldelli, P., & Meldolesi, J. (2015). The transcription repressor REST in adult neurons: physiology, pathology, and diseases. *eNeuro*, *2*, ENEURO.0010–15.2015. https://doi.org/10.1523/ENEURO.0010-15.2015.
37. Suo, H., Wang, P., Tong, J., Cai, L., Liu, J., Huang, D., … Huang, F. (2015). NRSF is an essential mediator for the neuroprotection of trichostatin

A in the MPTP mouse model of Parkinson's disease. *Neuropharmacology*, *99*, 67–78.
38. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., Van Der Lee, R., … Mathelier, A. (2018). JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, *46*, D260–D266.
39. Genuth, N. R., & Barna, M. (2018). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell*, *71*, 364–374.
40. Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature*, *381*, 571–580.
41. Zhang, X., Trépanier, V., Beaujois, R., Viranaicken, W., Drobetsky, E., & DesGroseillers, L. (2016). The downregulation of the RNA-binding protein Staufen2 in response to DNA damage promotes apoptosis. *Nucleic Acids Res.*, *44*, 3695–3712.
42. De Thonel, A., Le Mouël, A., & Mezger, V. (2012). Transcriptional regulation of small HSP – HSF1 and beyond. *Int. J. Biochem. Cell Biol.*, *44*, 1593–1612.
43. Li, J., Labbadia, J., & Morimoto, R. I. (2017). Rethinking HSF1 in stress, development, and organismal health. *Trends Cell Biol.*, *27*, 895–905.
44. Stephanou, A., & Latchman, D. S. (1999). Transcriptional regulation of the heat shock protein genes by STAT family transcription factors. *Gene Expr.*, *7*, 311–9.
45. Wang, Q., Li, W., Liu, X. S., Carroll, J. S., Jänne, O. A., Keeton, E. K., … Brown, M. (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol. Cell*, *27*, 380–392.
46. De Siervi, A., De Luca, P., Byun, J. S., Di, L. J., Fufa, T., Haggerty, C. M., … Gardner, K. (2010). Transcriptional autoregulation by BRCA1. *Cancer Res.*, *70*, 532–542.
47. Lu, X. (2010). Tied up in loops: Positive and negative autoregulation of p53. *Cold Spring Harbor Perspect. Biol.*, *2*, a000984. https://doi.org/10.1101/cshperspect.a000984.
48. Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C., & Asturias, F. J. (2008). Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol.*, *4*, 728–737.
49. Edwards, Y. J. K., Lobley, A. E., Pentony, M. M., & Jones, D. T. (2009). Insights into the regulation of intrinsically disordered proteins in the human proteome by analyzing sequence and gene expression data. *Genome Biol.*, *10*, R50.
50. Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.*, *13*, 227–232.
51. Edfors, F., Danielsson, F., Hallström, B. M., Käll, L., Lundberg, E., Pontén, F., … Uhlén, M. (2016). Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.*, *12*, 883.
52. Cirillo, D., Marchese, D., Agostini, F., Livi, C., Botta-Orfila, T., & Tartaglia, G. (2014). Constitutive patterns of gene expression regulated by RNA-binding proteins. *Genome Biol.*, *15*, R13.
53. Hansen, M. M. K., Wen, W. Y., Ingerman, E., Razooky, B. S., Thompson, C. E., Dar, R. D., … Weinberger, L. S. (2018). A Post-Transcriptional Feedback Mechanism for Noise Suppression and Fate Stabilization. *Cell*, *173*, 1609–1621.e15.e15.
54. Marchese, D., Botta-Orfila, T., Cirillo, D., Rodriguez, J. A., Livi, C. M., Fernandez-Santiago, R., … Tartaglia, G. G. (2017). Discovering the 3′ UTR-mediated regulation of alpha-synuclein. *Nucleic Acids Res.*, *45*, 12888–12903.
55. Pancaldi, V., & Bähler, J. (2011). In silico characterization and prediction of global protein-mRNA interactions in yeast. *Nucleic Acids Res.*, *39*, 5826–5836.
56. Zanzoni, A., Marchese, D., Agostini, F., Bolognesi, B., Cirillo, D., Botta-Orfila, M., … Tartaglia, G. G. (2013). Principles of self-organization in biological pathways: A hypothesis on the autogenous association of alpha-synuclein. *Nucleic Acids Res.*, *41*, 9987–9998.
57. Iserman, C., Roden, C., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., … Gladfelter, A. S. (2020). Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. *bioRxiv : the preprint server for biology*, 2020.06.11.147199.
58. Vandelli, A., Monti, M., Milanetti, E., Ponti, R. D., & Tartaglia, G. G. (2020). Structural analysis of SARS-CoV-2 and predictions of the human interactome. *Nucleic Acids Res.*, in press, https://doi.org/10.1101/2020.03.28.013789.
59. Seemann, S. E., Mirza, A. H., Hansen, C., Bang-Berthelsen, C. H., Garde, C., Christensen-Dalsgaard, M., … Gorodkin, J. (2017). The identification and functional annotation of RNA structures conserved in vertebrates. *Genome Res.*, *27*, 1371–1383.
60. Rivas, E., Clements, J., & Eddy, S. R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods*, *14*, 45–48.
61. Cerase, A., Armaos, A., Neumayer, C., Avner, P., Guttman, M., & Tartaglia, G. G. (2019). Phase separation drives X-chromosome inactivation: A hypothesis. *Nat. Struct. Mol. Biol.*, *26*, 331–334.
62. Ponti, R. D., Armaos, A., Marti, S., & Tartaglia, G. G. (2018). A method for RNA structure prediction shows evidence for structure in lncRNAs. *Front. Mol. Biosci.*, 5
63. Hershkovitz, E., Sapiro, G., Tannenbaum, A., & Williams, L. D. (2006). Statistical analysis of RNA backbone. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, *3*, 33–46.
64. Jolma, A., Zhang, J., Mondragón, E., Morgunova, E., Kivioja, T., Laverty, K. U., … Taipale, J. (2020). Binding specificities of human RNA-binding proteins toward structured and linear RNA sequences. *Genome Res.*, *30*, 962–973.
65. Ponti, R. D., Armaos, A., Vandelli, A., & Tartaglia, G. G. (2020). CROSSalive: A web server for predicting the in vivo structure of RNA molecules. *Bioinformatics*, *36*, 940–1.
66. Metkar, M., Ozadam, H., Lajoie, B. R., Imakaev, M., Mirny, L. A., Dekker, J., & Moore, M. J. (2018). Higher-order organization principles of pre-translational mRNPs. *Mol. Cell*, *72*, 715–726.e3.e3.
67. Tauber, D., Tauber, G., Khong, A., Van Treeck, B., Pelletier, J., & Parker, R. (2020). Modulation of RNA Condensation by the DEAD-Box Protein eIF4A. *Cell*, *180*, 411–426.e16.e16.
68. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., & Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, *505*, 701–705.
69. Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J - W., … Chang, H. Y. (2015). Structural imprints in vivo decode RNA regulatory mechanisms HHS Public Access. *Nature*, *519*, 486–490.
70. Alriquet, M., Calloni, G., Martínez-Limón, A., Ponti, R. D., Hanspach, G., Hengesbach, M., … Vabulas, R. M. (2020). The protective role of m1A during stress-induced granulation. *J. Mol. Cell Biol.*, https://doi.org/10.1093/jmcb/mjaa023 (in press).
71. Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., … Jaffrey, S. R. (2019). m6A enhances the phase separation potential of mRNA. *Nature*, *571*, 424–428.
72. Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.*, *30*, 39–58.
73. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.*, *24*, 329–332.
74. Lorenzo Gotor, N., Armaos, A., Calloni, G., Torrent Burgas, M., Vabulas, R. M., De Groot, N. S., & Tartaglia, G. G. (2020). RNA-binding and prion domains: The Yin and Yang of phase separation. *Nucleic Acids Res.*, *48*, 9491–9504.
75. Cerase, A., & Tartaglia, G. G. (2020). Long non-coding RNA-polycomb intimate rendezvous. *Open Biol.s*, *10*, 200126.
76. Ribeiro, D. M., Zanzoni, A., Cipriano, A., Delli Ponti, R., Spinelli, L., Ballarino, M., … Brun, C. (2018). Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. *Nucleic Acids Res.*, *46*, 917–928.

77. Mészáros, B., Erdos, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., … Pancsa, R. (2020). PhaSePro: The database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.*, *48*, D360–7.
78. You, K., Huang, Q., Yu, C., Shen, B., Sevilla, C., Shi, M., … Li, T. (2020). PhaSepDB: A database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.*, *48*, D354–D359.
79. Haify, S. N., Botta-Orfila, T., Hukema, R. K., & Tartaglia, G. G. (2020). In silico, in vitro, and in vivo Approaches to Identify Molecular Players in Fragile X Tremor and Ataxia Syndrome. *Front. Mol. Biosci.*, *7*, 31.
80. Palade, G. E. (1955). A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.*, *1*, 59–68.
81. Mitrea, D. M., & Kriwacki, R. W. (2016). Phase separation in biology; Functional organization of a higher order Short linear motifs – The unexplored frontier of the eukaryotic proteome. *Cell Commun. Signaling*, *14*, 1–20.
82. Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., … Brangwynne, C. P. (2016). Coexisting liquid phases underlie nucleolar subcompartments. *Cell*, *165*, 1686–1697.
83. Wolozin, B. (2012). Regulated protein aggregation: Stress granules and neurodegeneration. *Mol. Neurodegener.*, 7
84. Bolognesi, B., Gotor, N. L., Dhar, R., Cirillo, D., Baldrighi, M., Tartaglia, G. G., & Lehner, B. (2016). A concentration-dependent liquid phase separation can cause toxicity upon increased protein expression. *Cell Rep.*, *16*, 222–231.
85. Baldwin, A. J., Knowles, T. P. J., Tartaglia, G. G., Fitzpatrick, A. W., Devlin, G. L., Shammas, S. L., … Dobson, C. M. (2011). Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.*, *133*, 14160–14163.
86. Tartaglia, G. G., Cavalli, A., Pellarin, R., & Caflisch, A. (2005). Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci.*, *14*, 2723–2734.
87. Liu, Y., Fire, A. Z., Boyd, S., & Olshen, R. A. (2014). Estimating clonality. *Proc. Natl. Acad. Sci. USA*, *149*, 1048–59.
88. Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., … Alberti, S. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell*, *162*, 1066–1077.
89. Qamar, S., Wang, G. Z., Randle, S. J., Ruggeri, F. S., Varela, J. A., Lin, J. Q., … George-Hyslop, St, P. (2018). FUS Phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. *Cell*, *173*, 720–734.e15.e15.
90. Antonacci, G., de Turris, V., Rosa, A., & Ruocco, G. (2018). Background-deflection Brillouin microscopy reveals altered biomechanics of intracellular stress granules by ALS protein FUS. *Commun. Biol.*, *1*, 139.
91. Buratti, E. (2015). Functional significance of TDP-43 mutations in disease. *Adv. Genet.*, *91*, 1–53.
92. Ghosh, D. K., Kumar, A., & Ranjan, A. (2020). T54R mutation destabilizes the dimer of superoxide dismutase 1T54R by inducing steric clashes at the dimer interface. *RSC Adv.*, *10*, 10776–10788.
93. Chiti, F., & Dobson, C. M. (2017). Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.*, *86*, 27–68.
94. Tartaglia, G. G., Pechmann, S., Dobson, C. M., & Vendruscolo, M. (2009). A Relationship between mRNA Expression Levels and Protein Solubility in E. coli. *J. Mol. Biol.*, *388*, 381–389.
95. Agostini, F., Cirillo, D., Livi, C. M., Ponti, R. D., & Tartaglia, G. G. (2014). ccSOL omics: A webserver for large-scale prediction of endogenous and heterologous solubility in E. coli. *Bioinformatics*,*30*, 2975–2977.
96. Knowles, T. P. J., Vendruscolo, M., & Dobson, C. M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.*, *15*, 384–396.
97. Lee, Y., Zhou, T., Tartaglia, G. G., Vendruscolo, M., & Wilke, C. O. (2010). Translationally optimal codons associate with aggregation-prone sites in proteins. *Proteomics*, *10*, 4163–4171.
98. Yoo, H., Triandafillou, C., & Drummond, D. A. (2019). Cellular sensing by phase separation: Using the process, not just the products. *J. Biol. Chem.*, *294*, 7151–7159.
99. Taggart, J. C., & Li, G. W. (2018). Production of protein-complex components is stoichiometric and lacks general feedback regulation in eukaryotes. *Cell Systems*, *7*, 580–589.e4.e4.
100. Bergendahl, L. T., Gerasimavicius, L., Miles, J., Macdonald, L., Wells, J. N., Welburn, J. P. I., & Marsh, J. A. (2019). The role of protein complexes in human genetic disease. *Protein Sci.*, *28*, 1400–1411.
101. Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., … Vabulas, R. M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell*, *144*, 67–78.
102. Ciryam, P., Tartaglia, G. G., Morimoto, R. I., Dobson, C. M., & Vendruscolo, M. (2013). Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.*, *5*, 781–790.
103. Vecchi, G., Sormanni, P., Mannini, B., Vandelli, A., Tartaglia, G. G., Dobson, C. M., … Vendruscolo, M. (2020). Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc. Natl. Acad. Sci. USA*, *117*, 1015–1020.
104. Ciryam, P., Kundra, R., Morimoto, R. I., Dobson, C. M., & Vendruscolo, M. (2015). Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. *Trends Pharmacol. Sci.*, *36*, 72–77.
105. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., & Zweckstetter, M. (2017). Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.*, *8*, 1–13.
106. Chen, M., & Wolynes, P. G. (2017). Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. *Proc. Natl. Acad. Sci. USA*, *114*, 4406–4411.
107. Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., … Fuxreiter, M. (2018). Protein phase separation: A new phase in cell biology. *Trends Cell Biol.*, *28*, 420–435.
108. Markmiller, S., Soltanieh, S., Server, K. L., Mak, R., Jin, W., Fang, M. Y., … Yeo, G. W. (2018). Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell*, *172*, 590–604.e13.e13.
109. Rice, A. M., & McLysaght, A. (2017). Dosage-sensitive genes in evolution and disease. *BMC Biol.*, 15
110. Gsponer, J., & Babu, M. M. (2012). Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Rep.*, *2*, 1425–1437.
111. Sarge, K. D., Murphy, S. P., & Morimoto, R. I. (1993). Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. *Mol. Cell. Biol.*, *13*, 1392–1407.
112. Choi, S., Sa, M., Cho, N., Kim, K. K., & Park, S. H. (2019). Rbfox2 dissociation from stress granules suppresses cancer progression. *Exp. Mol. Med.*, 51
113. Gupta, N., Badeaux, M., Liu, Y., Naxerova, K., Sgroi, D., Munn, L. L., … Garkavtsev, I. (2017). Stress granule-associated protein G3BP2 regulates breast tumor initiation. *Proc. Natl. Acad. Sci. USA*, *114*, 1033–1038.
114. Quattrone, A., & Dassi, E. (2019). The architecture of the human RNA-binding protein regulatory network. *iScience*, *21*, 706–719.
115. Keene, J. D. (2007). RNA regulons: Coordination of post-transcriptional events. *Nat. Rev. Genet.*, *8*, 533–543.
116. Schaeffer, C. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.*, *20*, 4803–4813.
117. Ayala, Y. M., De Conti, L., Avendaño-Vázquez, S. E., Dhir, A., Romano, M., D'Ambrogio, A., … Baralle, F. E. (2011). TDP-43 regulates its

mRNA levels through a negative feedback loop. *EMBO J.*, *30*, 277–288.
118. Cirillo, D., Agostini, F., Klus, P., Marchese, D., Rodriguez, S., Bolognesi, B., & Tartaglia, G. G. (2013). Neurodegenerative diseases: Quantitative predictions of protein-RNA interactions. *RNA*, *19*, 129–140.
119. Polyansky, A. A., & Zagrovic, B. (2013). Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. *Nucleic Acids Res.*, *41*, 8434–8443.
120. Audic, Y., & Hartley, R. S. (2004). Post-transcriptional regulation in cancer. *Mol. Biol. Cell*, *96*, 479–498.
121. Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., & Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, *10*, e1004383.
122. Zacco, E., Martin, S. R., Thorogate, R., & Pastore, A. (2018). The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.*, 11
123. Manke, T., Bringas, R., & Vingron, M. (2003). Correlating protein-DNA and protein-protein interaction networks. *J. Mol. Biol.*, *333*, 75–85.
124. Klus, P., Cirillo, D., Botta Orfila, T., & Tartaglia, G. G. (2015). Neurodegeneration and cancer: Where the disorder prevails. *Sci. Rep.*, *5*, 15390.
125. Sanders, D. W., Kaufman, S. K., Holmes, B. B., & Diamond, M. I. (2016). Prions and protein assemblies that convey biological information in health and disease. *Neuron*, *89*, 433–448.
126. Santos, J., Iglesias, V., & Ventura, S. (2020). Computational prediction and redesign of aberrant protein oligomerization. In Progress in Molecular Biology and Translational Science. *Elsevier B.V.*, 43–83.
127. Kulkarni, U. (2019). Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules*, *9*, 147.
128. Hultqvist, G., Åberg, E., Camilloni, C., Sundell, G. N., Andersson, E., Dogan, J., … Jemth, P. (2017). Emergence and evolution of an interaction between intrinsically disordered proteins. *eLife*, *6*
129. Mann, J. R., Gleixner, A. M., Mauna, J. C., Gomes, E., DeChellis-Marks, M. R., Needham, P. G., … Donnelly, C. J. (2019). RNA binding antagonizes neurotoxic phase transitions of TDP-43. *Neuron*, *102*, 321–338.e8.e8.
130. Zacco, E., Graña-Montes, R., Martin, S. R., de Groot, N. S., Alfano, C., Tartaglia, G. G., & Pastore, A. (2019). RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *J. Mol. Biol.*, *431*, 1671–1688.
131. Batlle, C., de Groot, N. S., Iglesias, V., Navarro, S., & Ventura, S. (2017). Characterization of soft amyloid cores in human prion-like proteins. *Sci. Rep.*, *7*, 12134.
132. Tartaglia, G. G. (2016). The grand challenge of characterizing ribonucleoprotein networks. *Front. Mol. Biosci.*, *3*, 24.
133. Bellucci, M., Agostini, F., Masin, M., & Tartaglia, G. G. (2011). Predicting protein associations with long noncoding RNAs. *Nat. Methods*, *8*, 444–445.
134. Prusiner, S. B. (1998). Prions. *Proc. Natl. Acad. Sci. USA*, *95*, 13363–13383.
135. Soto, C., & Castilla, J. (2004). The controversial protein-only hypothesis of prion propagation. *Nat. Med.*, *10*, S63.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Armaos, A., Zacco, E., de Groot, N. S., & Tartaglia, G. G. (2021). RNA-protein Interactions: Central players in coordination of regulatory networks. *BioEssays*, 43, e2000118. https://doi.org/10.1002/bies.202000118