Systematic discovery of endogenous human ribonucleoprotein complexes

Anna L. Mallam[1,2,3,§,*], Wisath Sae-Lee[1,2,3], Jeffrey M. Schaub[1,2,3], Fan Tu[1,2,3], Anna Battenhouse[1,2,3], Yu Jin Jang[1], Jonghwan Kim[1], John B. Wallingford[1,2,3], Ilya J. Finkelstein[1,2,3], Edward M. Marcotte[1,2,3,§], Kevin Drew[1,2,3,§,*,†]

[1] Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA
[2] Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX 78712, USA
[3] Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA
[§] Correspondence: amallam@utexas.edu (A.L.M.), marcotte@icmb.utexas.edu (E.M.M.), kdrew@utexas.edu (K.D.)

* These authors contributed equally to this work

† Lead contact

Short title: A resource of human ribonucleoprotein complexes

Summary: An exploration of human protein complexes in the presence and absence of RNA reveals endogenous ribonucleoprotein complexes

Abstract

RNA-binding proteins (RBPs) play essential roles in biology and are frequently associated with human disease. While recent studies have systematically identified individual RBPs, their higher order assembly into Ribonucleoprotein (RNP) complexes has not been systematically investigated. Here, we describe a proteomics method for systematic identification of RNP complexes in human cells. We identify 1,428 protein complexes that associate with RNA, indicating that over 20% of known human protein complexes contain RNA. To explore the role of RNA in the assembly of each complex, we identify complexes that dissociate, change composition, or form stable protein-only complexes in the absence of RNA. Importantly, these data also provide specific novel insights into the function of well-studied protein complexes not previously known to associate with RNA, including replication factor C (RFC) and cytokinetic centralspindlin complex. Finally, we use our method to systematically identify cell-type specific RNA-associated proteins in mouse embryonic stem cells. We distribute these data as a resource, rna.MAP (rna.proteincomplexes.org) which provides a comprehensive dataset for the study of RNA-associated protein complexes. Our system thus provides a novel methodology for further explorations across human tissues and disease states, as well as throughout all domains of life.

Keywords: ribonucleoprotein complex, RNP, RNA binding protein, RBP, proteomics, DIF-FRAC, protein complexes, biochemical fractionation, mass spectrometry, interactome

Introduction

RNA-binding proteins (RBPs) play essential roles in diverse biological processes, and in most cases act within higher-order multi-protein complexes called Ribonucleoprotein (RNP) complexes (Castello et al., 2013; Gerstberger et al., 2014; Hentze et al., 2018). Understanding RNPs is of particular importance due to their indispensable role in many essential cellular functions, such as mRNA splicing (spliceosome) (Wahl et al., 2009), translation (ribosome) (Ramakrishnan, 2002), silencing (RISC) (Kawamata and Tomari, 2010), and degradation (exosome) (Houseley et al., 2006). Moreover, RNPs also play more specific roles in, for example, mRNA transport and localization in developing embryos and mature neurons (Holt and Bullock, 2009; Sahoo et al., 2018) and assembly phase separated organelles (Mittag and Parker, 2018). Further, RNPs are strongly implicated in human diseases including amyotrophic lateral sclerosis (ALS) (Scotter et al., 2015), spinocerebellar ataxia (Yue et al., 2001), and autism (Voineagu et al., 2011). Accordingly, substantial recent effort has been focused on systematic identification of RNA-associated proteins (Baltz et al., 2012; Bao et al., 2018; Brannan et al., 2016; Castello et al., 2012, 2016; He et al., 2016; Huang et al., 2018; Queiroz et al., 2019; Treiber et al., 2017; Trendel et al., 2019).

Strikingly, however, we still lack any systematic characterization of the assembly of individual RNA-associated proteins into the higher order RNP complexes in which so many function, leaving a crucial gap in our knowledge. A worldwide effort is currently underway to systematically identify multi-protein complexes using high throughput mass spectrometry techniques (Hein et al., 2015; Huttlin et al., 2015), but none of these techniques identify an RNA component within the complexes. We therefore set out to develop a method for systematic identification of RNA-associated higher-order multi-protein complexes that requires no genetic manipulation (i.e. tag-free) and would be easily adaptable to diverse cell types.

Here, we present 'differential fractionation for interaction analysis' (DIF-FRAC), which measures the sensitivity of protein complexes to a given treatment (ex. RNase A) using native size-exclusion chromatography followed by mass spectrometry. DIF-FRAC is based on a high throughput co-fractionation mass spectrometry (CF-MS) approach that we developed and applied to a diverse set of tissues and cells types en route to generating human and metazoan protein complex maps (Drew et al., 2017; Havugimana et al., 2012; Wan et al., 2015). DIF-FRAC builds upon CF-MS by comparing chromatographic separations of cellular lysate under control and RNA degrading conditions (Figure 1A). A statistical framework is then applied to discover RNP complexes by identifying concurrent shifts of known protein complex subunits upon RNA degradation (Figure 1A).

Analysis of DIF-FRAC data answers important questions as to the role of RNA plays in macromolecular complexes. Specifically, we identify RNP complexes that 1) dissociate, 2) form stable protein-only complexes, and 3) change composition in the absence of RNA suggesting specific roles for RNA in each of these cases. Because DIF-FRAC is independent of UV crosslinking, nucleotide incorporation, genetic manipulation or poly(A) RNA capture efficiency, it can therefore be used to investigate a wide variety of cell types, tissues and species. To demonstrate this versatility, we apply DIF-FRAC to mouse embryonic stem cells (mESCs), identifying 1,165 RNA-associated proteins, to show the method is highly adaptable and can be extended to discover RNP complexes in diverse samples.

Finally, we created a system-wide resource of 1,428 RNP complexes, many of which are previously unreported as having an RNA component, representing 20% of known human protein complexes. We provide our resource, rna.MAP, to the community as a fully searchable web database at rna.proteincomplexes.org.

Results and Discussion

Differential fractionation (DIF-FRAC) identifies RNP complexes

The DIF-FRAC strategy builds upon our previous strategy of Co-Fractionation Mass-spectroscopy (CF-MS) for identifying protein complexes in cellular lysate (Havugimana et al., 2012; Wan et al., 2015). CF-MS chromatographically separates protein complexes into fractions and uses a mass spectrometry pipeline to identify resident proteins in each fraction. The chromatographic elution profile of each protein is correlated to elution profiles from other proteins and similar profiles suggest physical interactions. Likewise, the DIF-FRAC strategy detects RNP complexes by identifying changes in the CF-MS elution profile of a protein complex's subunits upon degradation of RNA (Figure 1).

We applied DIF-FRAC to human HEK 293T cell lysate using size-exclusion chromatography (SEC) to separate the cellular proteins in a control and an RNase A-treated sample into 50 fractions (Figure 1A). Upon RNase A treatment, we observed a loss in the bulk chromatography absorbance signal in the high-molecular-weight regions and an increase in absorbance in lower molecular weight regions, consistent with higher-molecular-weight species (>1000 kDa) becoming lower-molecular-weight species in the absence of RNA (Figure 1B). The distribution of cellular RNA in these fractions measured using RNA-seq confirmed that we are accessing a diverse RNA landscape of mRNAs, small RNAs, and lncRNAs (Figure S1). As a negative control, we applied the same DIF-FRAC strategy to human erythrocytes, which we reasoned should have fewer RNPs since they have substantially lower amounts of RNA due to the loss of their nucleus and ribosomes upon maturation (Keerthivasan et al., 2011). Accordingly, the absorbance chromatography signal of erythrocyte lysate showed only a negligible difference in a DIF-FRAC experiment (Figure 1C). Together, these data establish that DIF-FRAC is capable of identifying bulk changes to the RNA-bound proteome.

We next used mass spectrometry to identify and quantify the resident proteins in each fraction for both the control and RNase A treated chromatographic separations, resulting in 8,946 protein identifications. Using these abundance measurements, we compared elution profiles (i.e. abundance change across chromatographically-separated molecular weights) between the control and RNase A treated experiments for each protein. A shift in a protein's elution profile between experiments is indicative of a protein-RNA interaction. For example, the known RNA helicase DDX21 shows a substantial shift in its elution profile upon RNase A treatment (Figure 1D), consistent with DDX21's known association with RNA (Calo et al., 2015). Alternatively, proteins such as the glucose synthesis enzyme, PGM1, show no shift, consistent with it not binding RNA (Figure 1E).

We can further examine these elution profile differences in the context of physically-associated proteins to identify RNP complexes. For example, subunits of the spliceosome, a known RNP complex, show elution profiles that coelute in the control but shift markedly upon RNA degradation (Figure 1F). In contrast, the elution profiles of subunits of the non-RNA-associated hexameric MCM complex (M_r ~550 kDa) (Figure 1G), as well as the 8-subunit COP9 signalosome (M_r ~500 kDa) (Figure S2A), are unchanged by RNase A treatment, consistent with the complexes not interacting with RNA. Thus, DIF-FRAC produces a robust signal that can be used to differentiate between non-RNA-associated complexes and RNP complexes.

Systematic identification of RNP complexes

In order to systematically identify RNP complexes in a DIF-FRAC experiment, we first developed a computational framework to identify statistically significant changes in individual proteins' elution behavior to identify RNA-associated proteins. We observed a variety of changes in elution behavior of known RNA-associated proteins upon RNase A treatment including decrease in molecular weight (e.g. NCL), increase in molecular weight (e.g. SUGP1), decrease in observed abundance (e.g. RPL18A), and increase in observed abundance (e.g. MACF1) (Figure 1H). To capture this range of behaviors in a simple metric, we developed the "DIF-FRAC score," which evaluates the degree to which two chromatographic separations differ (Figure 1I). Briefly, the DIF-FRAC score is a normalized Manhattan-distance between a protein's control and RNase A treated elution profiles (see Methods). To identify significant changes, we calculated P-values by comparing each protein's DIF-FRAC score to an abundance-controlled background distribution of DIF-FRAC scores from non-RNA-associated proteins (Figure S3A, see Methods for full description). We evaluated the score's performance on a curated set of known RNA-associated proteins and see strong correspondence between precision and high-ranking proteins (Figure 1J, Figure S3B). DIF-FRAC identifies 1012 proteins with significant elution profile differences in HEK 293T cells with a false discovery rate (FDR) cutoff of 5% (Table S1). To validate our metric, our set of statistically significant hits was compared to RNA-associated proteins identified from 11 other studies using alternative methods including RNA interactome capture (RIC) (Baltz et al., 2012; Castello et al., 2012), organic phase separation (Queiroz et al., 2019; Trendel et al., 2019), and others (Hentze et al., 2018). These results indicate that the DIF-FRAC score is highly accurate for identifying individual RNA-associated proteins (Figure S4A-K).

To expand on previous systematic studies of RNA-associated proteins, we exploited the unique features of DIF-FRAC to identify which RNA-associated proteins are assembled into higher order RNP complexes. Specifically, we searched for protein complexes whose subunits co-elute in the control experiment in addition to being sensitive to RNase A treatment (e.g. see Figure 1F). We detected 115 RNP complexes that fit these criteria, which we term 'RNP Select' (Figure 2 and Table S2)). The RNP Select set consists of 464 unique proteins, and importantly, it recapitulates many known RNP complexes. The set includes canonical RNPs such as the 40S ribosome (Figure S5) and the spliceosomal tri-snRNP complex, a major component of the catalytically active spliceosome that contains an intricate network of snRNA binding interactions (Agafonov et al., 2016) (Figure 2B). The set also includes RNPs with more specific functions such as the IGF2BP1 complex, which stabilizes c-myc RNA and prevents translation-coupled decay (Weidensdorfer et al., 2009) (Figure 2B). Because these data demonstrated the veracity of the DIF-FRAC strategy, we next searched our dataset for additional insights into RNP biology.

First, the RNP Select set provided new details about known RNPs. For example, stress granules are large membrane-less organelles that sequester mRNAs and prevent translation (Lin et al., 2015; Mittag and Parker, 2018; Spector and Lamond, 2011), and contain RNA-associated proteins including CAPRIN1, G3BP2, USP10 and NUFIP2 each localize to stress granules (Bardoni et al., 2003; Matsuki et al., 2013; Solomon et al., 2007). Interestingly, our previous map of human protein complexes (Drew et al., 2017) suggests that the known complex of G3BP, CAPRIN, and USP10 (Kedersha et al., 2016) also physically interacts with NUFIP2, leading us to suggest the name CapGUN (i.e. CAPRIN1, G3BP2, USP10, NUFIP2). Importantly, DIF-FRAC revealed that CapGUN subunits co-elute and associate with RNA (Figure 2B).

More importantly, RNP Select also contains several complexes not previously known to associate with RNA. For example, the spinal muscular atrophy associated activating signal cointegrator (ASC) complex (Knierim et al., 2016) (Figure 2B) is a transcriptional coactivator of nuclear receptors and has a role in transactivation of serum response factor (SRF), activating protein 1 (AP-1), and nuclear factor kappaB (NF-kappaB) (Jung et al., 2002). Upon RNase A treatment, we observed a substantial shift in elution from a high molecular weight to a lower molecular weight for all subunits of the ASC complex strongly suggesting that the complex associates with RNA (Figure 2). Interestingly, one ASC component, ASCC1, has a predicted RNA binding motif near its C terminus and has been shown to localize to nuclear speckles (Soll et al., 2018), which like stress granules are membraneless organelles enriched for RNPs. Our results, in coherence with previous studies point to a role for the ASC complex associating with RNA in RNP granules. Other notable examples of previously uncharacterized RNP complexes include the conserved oligomeric Golgi (COG) complex which is involved in intra-Golgi trafficking, and the SPATA5-SPATA5L1 complex, an uncharacterized complex linked to epilepsy, hearing loss, and mental retardation syndrome (Tanaka et al., 2015) (Figure 2B) among others (Table 1).

Finally, to ascertain the total number of annotated protein complexes that likely function with an RNA component, we evaluated DIF-FRAC evidence for RNA-associated proteins in addition to the 11 other studies (described above) and identify 1,428 complexes that contain a majority of RNA-associated proteins (see Methods). This analysis suggests that greater than 20% of known protein complexes associate with RNA (Table 1 and Table S2). We provide the complete set of RNP complexes as a fully searchable web database, rna.MAP, at http://rna.proteincomplexes.org. This represents a detailed resource of human RNP complexes, providing myriad testable hypotheses to guide further explorations of RNP biology.

Classification of RNP Complexes

RNA performs a variety of roles in macromolecular complexes. For example, it can bind as a substrate, function as an integral structural component, or act as a regulator of a complex's composition. Mirroring these roles, DIF-FRAC data reveals that upon RNA degradation, the proteins in RNP complexes can remain in an intact complex (Figure 3A), become destabilized (Figure 3B), or adopt different higher order configurations (Figure 3C). We therefore categorize RNP complexes into three groups.

The first category, which we term "apo-stable," defines protein complexes that remain stable after RNase A treatment. These include the exosome, RNase P, and the multi-synthetase complex (Figure 3A). Elution profiles of apo-stable complexes show that in the absence of RNA, subunits still co-elute, but do so as a lower-molecular weight complex. Available atomic structures of the exosome with and without RNA support the concept that RNA is peripheral to the stability of the complex (Gerlach et al., 2018; Weick et al., 2018).

The second category, which we designate as "structural", refers to complexes for which RNA is essential for the RNP complex structure and/or subunit solubility. These include, for example, the 60S and 40S ribosomal subcomplexes (Figure 3B and Figure S5). Upon degradation of RNA, the observed abundance of ribosomal protein subunits markedly decreases; suggesting the ribosome breaks apart and subunits become insoluble. This result is consistent with solved structures of the ribosome (Anger et al., 2013), demonstrating the centrality of rRNAs to the overall complex architecture (Figure 3B). Interesting exceptions to this behavior are the DIF-FRAC elution profiles for RPLP0, RPLP1 and RPLP2. These proteins co-elute in the RNase A treated sample, suggesting RNA does not mediate their interaction. Strikingly, however, this observation is consistent with the atomic structure of the human ribosome, which suggests that interactions between RPLP0, RPLP1 and RPLP2 are entirely protein-mediated (Figure 3B). This example demonstrates how DIF-FRAC data can not only identify RNA-protein mediated interactions, but can also provide structural information about RNP subcomplexes.

The third category, 'compositional' complexes, refers to those in which RNA promotes different stable combinations of protein-complex subunits, perhaps in a regulatory role (Figure 3C). For example, the WCRF (Williams syndrome transcription factor-related chromatin remodeling factor) complex, NuRD (Nucleosome Remodeling Deacetylase) complex and Cohesin complex are reported to assemble into a chromatin-remodeling supercomplex (CORUM ID: 282). We observed the WCRF and NuRD complexes co-eluting in the control experiment, forming a 12-subunit complex that shifts its elution upon RNA degradation. Interestingly, we also observed the supercomplex (WCRF, NuRD and Cohesin) eluting as a ~17-subunit complex in the RNA degradation condition. This composition change provides an explanation for why several NuRD-containing complexes are observed experimentally (Hakimi et al., 2002; Xue et al., 1998); our data suggest that these may represent both RNP complexes and non-RNA-associated complexes.

We also identified an uncharacterized compositional RNP complex containing the cell growth regulators DRG1 and ZC3H15 (DRFP1) (Ishikawa et al., 2005) that are implicated in lung cancer (Lu et al., 2016). ZC3H15 stabilizes DRG1 and prevents degradation possibly by preventing poly-ubiquitination (Ishikawa et al., 2005). Our result suggests that RNA is also involved in ZC3H15's role in stabilizing DRG1, as we observed a shift to a non-RNA-associated complex containing DRG1-ZC3H15 and LRRC41 in the absence of RNA (Figure 3C). LRRC41 is a probable substrate recognition component of E3 ubiquitin ligase complex (Kamura et al., 2004).

A further example of a compositional RNP complex is the transcription factor (TF)IIIC-TOP1-SUB1 complex, which is involved in RNA polymerase III pre-initiation complex (PIC) assembly (Male et al., 2015). DIF-FRAC shows this 7-subunit complex changes composition to the five-subunit TFIIIC upon RNA degradation (Figure 3C), offering further insights into the mechanism of TFIIIC-dependent PIC formation.

Finally, we identified the chromatin remodeling BRG/hBRM associated factors (BAF; the mammalian SWI/SNF complex; SWI/SNF-A) and polybromo-associated BAF (PBAF;

SWI/SNF-B) complexes as compositional RNP complexes, which is significant because these are some of the most frequently mutated protein complexes in cancer (Hodges et al., 2016; Tang et al., 2017) (Figure S6). BAF and PBAF complexes share a set of common core subunits, but also each have signature subunits that are related to their respective functions. Elution profiles show these core subunits co-elute with PBAF-only subunits in the control, but co-elute with BAF-only subunits upon RNA degradation (Figure S6). These data suggest BAF exists as a non-RNA-associated complex while PBAF functions as an RNP complex, consistent with its known role in transcription and supporting a previously described RNA-binding model where lncRNAs interact with SWI/SNF complexes in cancer (Tang et al., 2017). Together, these examples demonstrate the power of DIF-FRAC to describe the various physical relationships between RNA and macromolecular protein complexes.

Characterization of individual RNA-associated proteins

Although our efforts focused primarily on higher order RNP complexes, it is important to note that DIF-FRAC is also a powerful complement to existing methods for characterizing individual RNA-associated proteins. Indeed, DIF-FRAC identified 196 human RNA-associated proteins not previously identified in the many previous studies discussed in the Introduction, above (Table S3, Figure S4L). These DIF-FRAC identified RNA-associated proteins were strongly enriched in RNA binding domains annotated by Interpro (Finn et al., 2017) (Figure S4M). As we described above, inspection of elution profiles for the individual proteins revealed at least four distinct DIF-FRAC signals (Figure 1H, Figure 4). These manifest as elution-profile shifts with RNase A treatment that show: (1) an apparent decrease in molecular weight of the RNA-associated protein consistent with the degradation of an RNA component (Figure 4A); (2) a decrease in observed abundance, suggesting the RNA-associated protein becomes insoluble or is degraded (Figure 4B); (3) an apparent increase in molecular weight, suggesting the RNA-associated protein forms a higher-order species or aggregate (Figure 4C); or (4) an increase in observed abundance, indicative of the RNA-associated protein becoming more soluble (Figure 4D).

Analysis of all identified RNA-associated proteins shows 796 (79%) decrease in molecular weight, while 216 (21%) RNA-associated proteins increase in size (Figure S7). Aside from RNA acting as an interaction partner to RNA-associated proteins, RNA has been shown to regulate the oligomerization state of proteins both positively (Bleichert and Baserga, 2010; Huthoff et al., 2009; Xie et al., 2018) and negatively (Yoshida et al., 2004). Our data suggests that while the majority of RNA-associated proteins form higher-order assemblies with RNA, the oligomerization of 21% is potentially inhibited by RNA. Alternatively, RNA has also been shown to alter the solubility state of proteins (Maharana et al., 2018). We observe an increase in observed abundance for 535 (53%) proteins upon RNase A treatment, a decrease in abundance for 470 (47%) proteins, and no change in observed abundance for only 7 proteins. This suggests RNA impacts the solubility for most RNA-associated proteins and may function to tune protein availability in the cell.

Looking specifically at individual proteins provided insights that could impact our understanding of human disease. For example, we found that BANF1, a chromatin organizer, appears insoluble under our experimental conditions without RNA (Figure 4B). Interestingly, the BANF1 mutation Ala12-Thr12 causes Hutchinson-Gilford progeria syndrome, a severe and debilitating aging disease, by a reduction in protein levels (Puente et al., 2011). Our data suggest the hypothesis that this reduction is caused by disruption of the RNA-BANF1 interaction, leading to insolubility and degradation. Furthermore, RNA has also been shown to solubilize proteins linked to pathological aggregates (Maharana et al., 2018). Our data identifies a number of CREC family members (CALU, RCN1, RCN2, and SDF4; Figure 4C and Table S1) as RNA-associated proteins that increase in molecular weight upon RNA degradation. The CREC family is a group of multiple EF-hand, low-affinity calcium-binding proteins with links to amyloidosis (Vorum et al., 2000). DIF-FRAC demonstrates a dependence of RNA on the oligomerization state of CALU, which could play a role in the formation of amyloid deposits similar to that observed for prion-like RNA-associated proteins (Maharana et al., 2018). Based on these examples and the many disease links to DIF-FRAC identified RNP complexes (Table 1), we anticipate our data will generate testable RNA-related hypotheses about disease-related states.

Directed validation of Replication factor C (RFC) as an RNP Complex

An important aspect of DIF-FRAC is that while it provides a systematic survey, the experimental basis for each data point can be directly assessed in the elution profiles. Nonetheless, the ultimate demonstration of the utility of any large-scale dataset is its ability to make predictions that can be validated by orthogonal experiments. Among the most surprising findings in our data was that the extensively-characterized replication factor C (RFC) complex (Yao and O'Donnell, 2012) exists as a stable RNP complex (Figure 2A). During replication and DNA damage repair, the RFC complex is responsible for loading PCNA, a DNA polymerase processing factor, onto DNA. Strikingly, DIF-FRAC identified two previously observed variants of the RFC complex, RFC1-5 and RFC2-5 (Figure 5A), and more importantly demonstrated that RFC1-5 appears to be the dominant variant and is also the RNA-associated form (Figure 5C). Consistent with the RFC complex interacting with RNA, the homologous clamp loader in *E.coli*, γ complex, is known to load the DNA clamp onto RNA-primed template DNA (Yao and O'Donnell, 2012) and eukaryotic RFC has also been shown to be capable of loading PCNA onto synthetic RNA-primed DNA (Yuzhakov et al., 1999). In light of this finding, we tested whether purified RFC complex from *S. cerevisiae* could directly bind different species of nucleic acids. We observed that RFC not only binds dsDNA and DNA-RNA hybrids, but also binds dsRNA with surprisingly tight binding constants in the nanomolar range (Figure 5D and Figure S8). These data show RFC binds dsRNA and point to an uncharacterized role for RNA in the function of RFC. The result also further validates the use of DIF-FRAC to identify uncharacterized RNP complexes.

Evaluating RNPs in multiple proteomes

Finally, because DIF-FRAC does not rely on any specialized reagents, the strategy can be applied to any cell type that can be readily isolated. Because of the long-standing interest in the role of RNPs in embryonic development (e.g. for targeted localization of maternal RNAs (Escobar-Aguirre et al., 2017); processing of non-coding RNA to direct differentiation and stem cell potency (Dinger et al., 2008; Guttman et al., 2011; Yan et al., 2013)), we applied DIF-FRAC to mouse embryonic stem cells (mESCs). We identified 1,165 significant RNA-associated proteins in mESCs (Figure 6A, Table S1), including 466 that are novel, representing a 35% increase in the number of annotated mouse RNA-associated proteins (Figure 6B). This mESC dataset provides three advances:

First, the data can provide additional evidence to support assignment of novel RNPs. For example, many of the RNA-associated proteins identified in mESCs reflected equivalent RNA-associated proteins in human cells (Figure 6C), including the RFC complex, which specifically behaves as a compositional RNP complex in both species (Figure 5B).

Second, this approach allowed the identification of cell-type specific RNA-associated proteins. Indeed, we identified several mESC-specific RNA-associated proteins and these included several that have been previously implicated in stem cell function. For example, we identified the known pluripotency factor Sox2 (Figure 6D) and the Polycomb Repressor complex 2 subunit Jarid2 (Figure 6E), as RNA-associated proteins, consistent with previous reports (Cifuentes-Rojas et al., 2014; Fang et al., 2011; Kaneko et al., 2014).

Finally, the additional dataset allows us to cast a wider net in our search for novel RNPs. For example, among the RNA-associated proteins identified in mESCs were members of the centralspindlin complex, a heterotetramer consisting of Racgap1 and Kif23 and involved in cytokinesis (White and Glotzer, 2012; Yüce et al., 2005). Previously unknown to contain an RNA component, we identify Racgap1, Kif23 and the centralspindlin interaction partner Ect2 as significantly sensitive to RNase A treatment in mESCs (Figure 6F). In agreement with this mESC result, we observed a similar trend for this complex in human cells, showing conservation across species (Figure 6G). Our results suggest a physical interaction between the centralspindlin complex and RNA, thus informing a previous study that report Kif23 (ZEN-4 in *C.elegans*) as a positive regulator of RNP granule formation (Wood et al., 2016), as well as the localization of several RNA species to the midbody during cytokinesis (Clemson et al., 1996; Lécuyer et al., 2007; Zheng et al., 2010).

Together, these data demonstrate that the adaptability of DIF-FRAC to diverse systems will allow identification of conserved RNA-associated proteins and RNP complexes in diverse tissues and disease states across all domains of life.

Conclusion

Here, we report the design, development, and application of a robust fractionation-based strategy to determine RNP complexes on a proteome-wide scale. We successfully used DIF-FRAC to identify 115 stable RNP complexes throughout the human interactome, and applied it to multiple tissue types and species. Combining this with previous data, we generate a resource of the RNA-bound human proteome and demonstrate that upwards of 20% of protein complexes contain an RNA component, highlighting the prevalence of RNP complexes in the cellular milieu. Together our results provide a valuable tool for researchers to investigate the role of RNPs in protein function and disease.

The DIF-FRAC methodology offers important advances over previous techniques that examine RNA-protein interactions. Specifically, interactions are probed proteome-wide in a native, whole lysate sample using a strategy that is not reliant on labeling or cross-linking efficiency. We show DIF-FRAC can be applied effectively to multiple cell types and organisms, and has the potential to provide information on protein-RNA interactions in disease states. Furthermore, DIF-FRAC is a broadly applicable framework that can be extended to examine other large-scale proteomic changes in a system of interest.

We also introduce three classifications of RNP complexes (apo-stable, structural, and compositional) that provide a useful framework to organize the roles of RNAs in macromolecular complexes. Additionally, DIF-FRAC provides information on the biochemical characteristics (i.e. molecular weight, solubility) of RNP complexes in the presence and absence of RNA that offer clues to disease pathophysiology. We anticipate this technique to be a powerful tool to uncover the molecular mechanisms of RNA related diseases. Overall, the DIF-FRAC method described and demonstrated here charters new territories in the cellular landscape of RNA-protein interactions. We have utilized DIF-FRAC to provide the first system-wide resource of human RNPs, providing a broadly applicable tool for studying cellular interactions and responses in multiple cell types and states.

Acknowledgements

We would like to thank Matt Davis for helpful comments, Claire D. McWhite for plotting software and the Texas Advanced Computing Center for high-performance computing. This work was supported by grants from the NIH (K99 HD092613 to K.D., R35 GM122480, DP1 GM106408, R21 GM119021, R01 DK110520, and R01 HD085901 to E.M.M., R01 GM112722 to J.K. and R01 GM120554 to I.J.F.); NSF (IOS-1237975 to E.M.M., and 1453358 to I.J.F.); Army Research Office (W911NF-12-1-0390 to E.M.M.); and The Welch Foundation (F1515 to EMM and F1808 to I.J.F.). Data collection by the UT Austin Proteomics Facility was supported by CPRIT RP110782 to Maria Person.

Figure Legends:

Figure 1: Differential fractionation (DIF-FRAC) identifies RNP complexes. (A) The DIF-FRAC workflow requires two equivalent cell culture lysates for a control and an RNase A treated sample. Lysate is separated into fractions by size exclusion chromatography (SEC), and proteins in each fraction are identified by mass spectrometry to determine individual protein elution profiles proteome-wide for each condition. An elution shift of a protein upon RNase A treatment is indicative of an RNA-protein association. Elution shifts are cross-referenced with known protein complexes to identify RNP complexes. (B) Separations of HEK293T lysate under control (black) and RNase A treated (red) conditions monitored by bulk SEC chromatography absorbance profiles at A_{280} show loss of high molecular weight signal upon treatment. (C) Negative control separations of erythrocyte lysate under control (black) and RNase A treated (red) conditions monitored by bulk SEC chromatography absorbance profiles at A_{280} show no change in absorbance signal. (D) RNA binding protein elution profile for positive control nucleolar RNA helicase 2 (DDX21) (Abundance = count of unique peptide spectral matches). The elution profile shows sensitivity to RNase A treatment. (E) Elution profile for negative control phosphoglucomutase (PGM1) is not sensitive to RNase A treatment. (F) Elution profiles for subunits of the spliceosome RNP complex (i.e. positive control) show co-elution of complex in control and a shift in elution upon RNase A treatment. (G) Elution profile for the non-RNA-associated MCM complex (i.e. negative control) shows no detectable elution shift. (H) Example traces of four known RNA-associated proteins exhibiting different behaviors of elution profile changes upon RNase A treatment. NCL shows a loss in molecular weight while SUGP1 shows an increase in molecular weight. RPL18A shows a decrease in observed abundance while MACF1 shows an increase in observed abundance. In (B)-(H) dashed lines correspond to the elution volumes of molecular weight standards thyroglobulin (M_r = 669 kDa), apoferritin (M_r = 443 kDa), albumin (M_r = 66 kDa), and carbonic anhydrase (M_r = 29 kDa). Molecular weight labels on subsequent plots are removed for clarity. (I) A DIF-FRAC score is calculated for each protein from the absolute value of the difference of the elution profiles between control and RNase A treated samples, and then summed. A P-value is then calculated from a background distribution of DIF-FRAC scores. See also Figure S3A. (J) DIF-FRAC P-value calculated on HEK 293T data shows strong ability to discriminate known RNA-associated proteins from other proteins. See also Figure S3B.

Figure 2: DIF-FRAC reveals a map of stable RNP complexes. (A) 115 RNP complexes identified by the DIF-FRAC method termed 'RNP Select'. Green nodes represent proteins annotated as 'ribonucleoprotein complex' and yellow nodes are unannotated proteins. RNP Select complexes are defined as complexes whose protein subunits co-elute in the control DIF-FRAC sample (> 0.75 average correlation coefficient) and > 50% of subunits have a DIF-FRAC P-value > 0.5. DIF-FRAC identified many known RNP complexes such as the ribosome, mitochondrial ribosome and snRNP as well as novel RNP complexes such as RFC, COG, ASC and SPATA5. (B) Individual RNP complexes with elution profiles including *(i)* snRNP, *(ii)* IGF2BP1, *(iii)* CapGUN, *(iv)* COG, *(v)* ASC, and *(vi)* SPATA5. Abundance represents count of unique peptide spectral matches for each protein.

Figure 3: DIF-FRAC identifies 3 classes of RNP complexes. (A) 'Apo-stable' RNP complexes: elution profiles of the exosome (top, CORUM 7443), RNase P (middle, CORUM 123), and the multi-synthetase complex (bottom, CORUM 3040) show each complex is a stable complex that binds RNA and the complex remains intact in the absence of RNA. Blue shading represents RNA bound complex. Blue shading represents RNA bound form and red shading represents RNA unbound complex. (B) 'Structural' RNP complexes: elution profiles of the 60S ribosomal subunit (CORUM 308) show the complex destabilizes upon RNA degradation and subunits no longer co-elute upon RNase A treatment. DIF-FRAC elution data show the ribosomal subunits RPLP0, RPLP1 and RPLP2 (orange) remain as a subcomplex upon RNA degradation, consistent with their position in the solved ribosome structure whose interactions are not mediated by RNA (bottom, PDB 4V6X, protein in blue, RNA in red, ribosomal stalk in orange). (C) 'Compositional' RNP complexes: (Top) Elution profiles of WCRF-Cohesin-NuRD (CORUM 282) and NuRD-WCRF suggest that RNA association promotes different forms of the complex. (Middle) Elution profiles of Drg1-ZC3H15-LRRC41 complex (hu.MAP 2767) which forms only in the absence of RNA. (Bottom) Elution profiles of the TFIIIC containing-TOP1-SUB1 complex (CORUM 1106) loses two subunits, TOP1 and SUB1, upon RNA degradation. Green shading represents RNA unbound complex. In (A)-(C) vertical dashed lines correspond molecular weight standards described in Figure 1.

Figure 4: DIF-FRAC identifies four distinct signals for RNA-associated proteins. Examples of elution profiles for disease related proteins that (A) decrease in size, MAP1A, (B) decrease in observed abundance (less soluble), BANF1, (C) increase in size, RCN1, and (D) increase in observed abundance (more soluble), HMMR, upon RNA degradation.

Figure 5: Replication factor C is an RNP Complex. Elution profiles in both human (A) and mouse (B) demonstrates RFC1-5 forms an RNP complex (blue/yellow highlight). A smaller subcomplex of RFC2-5 (green highlight) becomes the dominant form upon RNA degradation. (C) A cartoon to show the RNA-dependence of annotated complexes RFC1-5 (blue) and RFC2-5 (green) as determined by DIF-FRAC. RNA is shown in grey. (D) Electromorphic mobility shift assays (EMSA) of various concentrations of purified *S. cerevisiae* RFC mixed with 1 nM ^{32}P-labeled oligonucleotides. Representative gels show RFC binds dsDNA, DNA/RNA hybrid and dsRNA substrates. RFC-nucleic acid complexes were separated on 10% native gels. Binding constants are in the nanomolar range (Figure S8).

Figure 6: DIF-FRAC identifies RNP complexes across cell types and species. (A) DIF-FRAC identifies 1,165 RNA-associated proteins in mESC (mouse embryonic stem cells), and 1,012 RNA-associated proteins in HEK 293T cells. (B) Venn diagram of considerable overlap between previously published large-scale RNA-protein interaction studies, literature annotated RNA-associated proteins, and DIF-FRAC identified RNA-associated proteins in mESC. (C) RNA-associated human-mouse orthologues are identified reproducibly in DIF-FRAC experiments. Elution profiles for known pluripotency factors, Sox2 (D) and Jarid2 (E) show association with RNA in mESC cells. Elution profiles of the centralspindlin complex for (F) mESC and (G) HEK 293T demonstrate centralspindlin is an RNP complex in both species. Yellow and blue shading represent RNA bound complex in mESC and HEK 293T respectively.

Figure S1: DIF-FRAC accesses a diverse RNA landscape. Box plots show the RNA abundance of mRNA, lncRNA, small RNA, other ncRNA, and pseudogenes in control fractions 16-23 of HEK 293T cell lysate (TPM = Transcripts Per Million). Boxes indicate median (inner joint), first quartile (left) and third quartile (right). Lines indicate 1.5 interquartile range. Dots indicate outliers.

Figure S2: Non-RNA-associated complexes are insensitive to RNase A treatment. DIF-FRAC elution profiles show subunits of the negative control non-RNA-associated COP9 signalosome complex (M_r ~500 kDa (Oron et al., 2002)) in control (black) and RNase A treated (red) for (A) HEK293T lysate, and (B) mESC do not shift upon RNase A treatment. Abundance represents count of unique peptide spectral matches. Vertical dotted lines represent protein standards described in Figure 1.

Figure S3: DIF-FRAC Score accurately discriminates between RNA-associated proteins and non-binders.

(A) Workflow to calculate abundance corrected P-values for each protein's DIF-FRAC score. Proteins are ranked according to abundance and a window of +/- 100 proteins is used to calculate a DIF-FRAC score distribution. A two-component Gaussian mixture model is then used to identify the non-RNA binding component in the distribution. Finally, the DIF-FRAC score of the protein of interest is compared to the non-RNA binding distribution component to test the null hypothesis and a P-value is calculated. (B) Precision recall analysis shows the DIF-FRAC Score recalls a substantial number of known RNA-associated proteins in HEK 293T cells. (C) High DIF-FRAC P-values have high precision in recovering known RNA-associated proteins in mouse embryonic stem cells. (D) Precision recall analysis shows the DIF-FRAC Score recalls a substantial number of known RNA-associated proteins in mESC.

Figure S4: DIF-FRAC RNA-associated proteins show substantial overlap with other high-throughput studies. (A-K) Venn diagrams show overlap of DIF-FRAC RNA-associated proteins from HEK 293T cells (blue) with 11 high-throughput RNA association studies (green) (Bao et al., 2018; Hentze et al., 2018; Huang et al., 2018; Queiroz et al., 2019; Trendel et al., 2019). (L) Venn diagram shows overlap of DIF-FRAC RNA-associated proteins from HEK 293T cells (blue), annotated RNA Binding proteins (red), and combined set of high throughput RNA association studies (green). (M) Enrichment of RNA binding structural motifs in DIF-FRAC-identified RNA-associated proteins from HEK 293T cells.

Figure S5: DIF-FRAC classifies the 40S ribosomal subunit as a structural RNP. Elution profiles of the 40S ribosomal subunits demonstrate it is destabilized upon RNA degradation ('structural' RNP complex).

Figure S6: DIF-FRAC classifies human BAF and PBAF complexes as compositional RNPs. Elution profiles of annotated human PBAF (blue shading) and BAF (green shading) complexes demonstrate core subunits common to both complexes coelute in both control and RNase A treated samples, but at different molecular weights. PBAF-only subunits (light grey) coelute with core subunits only in the control sample, while BAF-only subunits (dark grey) coelute with the core subunits as a lower molecular weight complex only when RNA is degraded. Together, these elution profiles suggest that PBAF is an RNP complex, but the BAF complex does not associate with RNA.

Figure S7: Analysis of DIF-FRAC shift types of RNA-associated proteins. Upon RNase A treatment we observe different types of changes to protein elution profiles (see Figure 1H and Figure 4). Each point in the graph represents one RNA-associated protein. Molecular weight shift is the weighted average difference between control and RNase A treated profiles, where a negative value (left side of graph) represents lower molecular weight elution upon treatment and positive value (right side of graph) represents gain in molecular weight (see Methods for calculation). Abundance change is the normalized change in observed abundance upon RNase A treatment. A positive value (top of graph) represents gain in solubility and a negative value (bottom of graph) represents loss in solubility. Examples from Figure 4 are annotated.

Figure S8: Affinity of nucleic acid for the *S. cerevisiae* RFC complex. Binding curves from electromorphic mobility shift assays (EMSA) of various concentrations of purified *S. cerevisiae* RFC mixed with 1 nM ^{32}P-labeled oligonucleotides. Data was fit to a hyperbolic equation (solid line). The calculated k_D + 95% CI is 1.9 ± 0.5 nM for dsDNA (black), 7.5 ± 4.9 nM for DNA/RNA hybrid (red), and 25 ± 11 nM for dsRNA (blue). Error bars denote standard deviation.

Materials and Methods

Experimental Design

Human cell culture and extract preparation

HEK293T cells (ATCC CRL3216) cultured in DMEM (Gibco) supplemented with 10 % (v/v) FBS (Life Technologies) were continually split over 7 days to give four 10-cm dishes of adherent cells. For the control fractionation sample, two 10-cm dishes of cells were harvested at 80-100 % confluence without trypsin by washing in ice cold phosphate buffered saline (PBS) pH 7.2 (0.75 mL; Gibco) and placed on ice. Cells (approximately 0.1 g wet weight) were lysed on ice (5 min) by resuspension in Pierce IP Lysis Buffer (0.8 mL; 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40 and 5% glycerol; Thermo Fisher) containing 1x protease inhibitor cocktail III (Calbiochem). The resulting lysate was clarified (17,000g, 10 min, 4°C) and left at room temperature (30 min). The sample was filtered (Ultrafree-MC filter unit (Millipore); 12,000g, 2 min, 4°C) to remove insoluble aggregates. RNase A treated samples were prepared on the same day in an identical manner, except RNase A (8 μL, 80 μg, Thermo Fisher, catalogue #EN0531) was added after lysate clarification and the sample left at room temperature (30 min) before filtration.

Mouse embryonic stem cell culture

Gelatin adapted mouse J1 ES cells (ATCC® SCRC-1010™) were cultured in Dulbecco’s Modified Eagle’s Medium (DMEM, Life Technologies) containing 18% fetal bovine serum (FBS, Gemini), 50 U/mL of penicillin/streptomycin with 2 mM L-glutamine (Life Technologies), 0.1 mM non-essential amino acid (Life Technologies), 1% nucleosides (Sigma-Aldrich), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), and 1,000 U/mL recombinant leukemia inhibitory factor (LIF, Chemicon). ES cells were plated on 15-cm dishes coated with 0.1% gelatin and incubated at 37°C and 5% CO_2. Cells were passaged every 2 days. Lysis and RNase A treatment were done as described in the HEK 293T protocol.

Erythrocyte cell preparation

Leukocyte-reduced red blood cells (RBCs) were obtained from an anonymous donor and purchased from Gulf Coast Regional Blood Center (Houston, Texas). The RBCs used in this experiment were kept at 4°C for 54 days before lysis to ensure reticulocytes mature into RBCs.

Prior to cell lysis, RBCs were washed with ice cold PBS (pH 7.4, Gibco) for 3 times at 600 g for 15 min at 4°C. RBCs were then lysed in hypotonic solution (5 mM Tris-HCl, pH 7.4) containing protease and phosphatase inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail, Roche and PhosSTOP, Roche) with a ratio of 1 volume packed RBC: 5 volumes hypotonic solution. Hemolysate (soluble fraction of RBC lysate) was collected by centrifuging white ghosts (membrane fraction of RBC lysate) at 21,000 g for 40 mins at 4°C. Hemolysate was collected and stored at -80°C until further use. On the day of experiment, hemolysate was thawed and treated with Hemoglobind (Biotech Support Group) in order to remove hemoglobin from hemolysate. A total of 4-5 mg of total proteins were split into control and RNase A treated samples. The RNase sample was treated with RNase A as described in the protocol of RNase A treatment of lysate from HEK293T cells. Both samples were filtered (Ultrafree-MC filter unit (Millipore); 12,000 g, 2 min, 4°C) to remove insoluble aggregates prior to fractionation.

Biochemical fractionation using native size-exclusion chromatography

All lysates were subject to size exclusion chromatography (SEC) using an Agilent 1100 HPLC system (Agilent Technologies, ON, Canada) with a multi-phase chromatography protocol as previously described (Havugimana et al., 2012). Soluble protein (1.25 mg, 250 µL) was applied to a BioSep-SEC-s4000 gel filtration column (Phenomenex) equilibrated in PBS, pH 7.2 (HEK 293T and mESC lysate) or pH 7.4 (erythrocytes) at a flow rate of 0.5 mL min^{-1}. Fractions were collected every 0.375 mL. The elution volume of molecular weight standards (thyroglobulin (M_r = 669 kDa); apoferritin (M_r = 443 kDa); albumin (M_r = 66 kDa); and carbonic anhydrase (M_r = 29 kDa); Sigma) was additionally measured to calibrate the column (Figure 1B).

Mass spectrometry

Fractions were filter concentrated to 50 µL, denatured and reduced in 50 % 2,2,2-trifluoroethanol (TFE) and 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 55 °C for 45 minutes, and alkylated in the dark with iodoacetamide (55 mM, 30 min, RT). Samples were diluted to 5 % TFE in 50 mM Tris-HCl, pH 8.0, 2 mM $CaCl_2$, and digested with trypsin (1:50; proteomics grade; 5 h; 37 °C). Digestion was quenched (1 % formic acid), and the sample volume reduced to ~100 µL by speed vacuum centrifugation. The sample was washed on a HyperSep C18 SpinTip (Thermo

Fisher), eluted, reduced to near dryness by speed vacuum centrifugation, and resuspended in 5 % acetonitrile/ 0.1 % formic acid for analysis by LC-MS/MS.

Peptides were separated on a 75 μM x 25 cm Acclaim PepMap100 C-18 column (Thermo) using a 3-45 % acetonitrile gradient over 60 min and analyzed online by nanoelectrospray-ionization tandem mass spectrometry on an Orbitrap Fusion or Orbitrap Fusion Lumos Tribrid (Thermo Scientific). Data-dependent acquisition was activated, with parent ion (MS1) scans collected at high resolution (120,000). Ions with charge 1 were selected for collision-induced dissociation fragmentation spectrum acquisition (MS2) in the ion trap, using a Top Speed acquisition time of 3-s. Dynamic exclusion was activated, with a 60-s exclusion time for ions selected more than once. MS from HEK 293T cells was acquired in the UT Austin Proteomics Facility.

Construction and sequencing of RNA-seq libraries of DIF-FRAC samples

Fractions from a biological replicate SEC separation corresponding to higher molecular weight species (approximately >1.5 MDa; fractions 16-23 in Figure 1B) were analyzed by total RNA sequencing. Total RNA was isolated from each fraction (0.375 mL) by addition of Trizol (1.125 mL; Thermo Fisher) and the sample (1.4 mL) was transferred to a Phasemaker tube (Thermo Fisher). Total RNA was extracted following the protocol supplied by the manufacturer and further cleaned up using a RNeasy MinElute Cleanup Kit (Qiagen). RNA integrity number (RIN) was measured using an Agilent Bioanalyzer and samples were ribo-depleted using a using a RiboZero Gold (Human/Mouse/Rat) kit (Illumina) to remove rRNAs. RNA libraries were prepared for sequencing according to vendor protocols using NEBNext R Small RNA Library Prep Set for Illumina R (Multiplex Compatible), Cat #E7330L, according to the protocol described by Podnar et al. (Podnar et al., 2014). RNA was fragmented using elevated temperature in carefully controlled buffer conditions to yield average fragment sizes of 200 nucleotides. These fragments were directionally ligated to 5′ and 3′ adaptors so that sequence orientation is preserved throughout sequencing. Reverse transcription and PCR were performed to complete the DNA sequencing libraries, which were sequenced using an Illumina NextSeq 500 instrument (75-nt single reads) at the Genomic Sequencing and Analysis Facility at the University of Texas at Austin.

S. cerevisiae RFC purification:

RFC was purified as previously described (Finkelstein et al., 2003; Kim et al., 2017). Briefly, full-length *S. cerevisiae* RFC was expressed in BL21(DE3) ArcticExpress (Agilent) *E. coli* co-transformed with pLant2b-RFC-AE (pIF117) and pET11-RFC-BCD (pIF116). RFC was subsequently purified by SP and Q (GE Healthcare) ion exchange chromatography. Protein concentration was determined by comparison to a BSA titration curve using Coomassie-stained SDS-PAGE.

Electrophoretic Mobility Shift Assay (EMSA):

Oligonucleotide constructs were based on an earlier description (Kobayashi et al., 2006). Each of the four-nucleic acid substrates were radiolabeled with [γ-^{32}P]-ATP using T4 Polynucleotide Kinase (NEB). Free nucleotide was removed using G-25 MicroSpin columns (GE Healthcare). Oligonucleotides were subsequently heated to 75°C and slowly cooled to room temperature to allow proper annealing. 1 nM oligonucleotide and various concentrations of RFC (0 to 256 nM) were incubated for 15 minutes at room temperature in a buffer containing 25 mM Tris-HCl [pH 7.5], 50 mM NaCl, 2 mM $MgCl_2$, 2 mM DTT, and 0.1 mg/mL BSA. Reactions were quenched with 6x loading dye (10 mM Tris-HCl [pH 7.6], 60% glycerol, 60 mM EDTA, 0.15% [w/v] Orange G) and subsequently separated by native acrylamide gel electrophoresis. Gels were dried on Zeta-Probe Membrane (Bio-Rad) at 80°C for two hours. Bands were visualized by a Typhoon FLA 7000 phosphorimager (GE Healthcare). Binding was quantified using FIJI (Schindelin et al., 2012). Subsequent data were fit to a hyperbolic equation to determine the k_D for oligonucleotide binding.

Oligonucleotides used:

Name	Sequence
dsDNA	5’ - CTC GAG GTC GTC ATC GAC CTC GAG ATC A – 3’
DNA/RNA	5’ - rCrUrC rGrArG rGrUrC rGTC ATC GAC CTC GAG ATC A – 3’
dsRNA	5’ - rCrUrC rGrArG rGrUrC rGrUrC rArUrC rGrArC rCrUrC rGrArG rArUrC rA – 3’

Calculated k_D from fitting to hyperbolic equation (Bound = (v*[E])/(k_D+[E])), where “[E]” is the concentration of the enzyme, and “v” and “k_D” are solved by linear regression.

Bioinformatic analysis

Protein identification

Prior to protein identification, human and mouse proteomes were downloaded from UniProt website (Apweiler et al., 2004). Raw formatted mass spectrometry files were first converted to mzXML file format using MSConvert (http://proteowizard.sourceforge.net/tools.shtml) and then processed using MSGF+ (Kim et al., 2017), X! TANDEM (Craig and Beavis, 2004) and Comet (Lingner et al., 2011) peptide search engines with default settings. MSBlender (Kwon et al., 2011) was used for integration of peptide identifications and subsequent mapping to protein identifications. A false discovery rate of 1% was used for peptide identification. Protein elution profiles were assembled using unique peptide spectral matches for each protein across all fractions collected.

DIF-FRAC score and P-value significance calculation

In order to determine the significance of a protein's sensitivity to RNase A treatment, we compare the protein's control elution profile to its RNase A treated elution profile as schematized in Figures 2A and S3A. Specifically, we first calculate the L1-norm of the two elution profiles (**equation 1**).

$$D_p = \sum_{i=1}^{N} \mid X_{p,i} - Y_{p,i} \mid \tag{1}$$

Where *N* represents the total number of fractions collected and *p* represents an individual protein. *X* and *Y* represent abundance matrices of control and experiment (RNase A treated) respectively. We next normalize D_p by the total abundance seen for protein *p* in both the control and experiment (**equation 2**).

$$D_p^{norm} = \frac{D_p^2}{\sum_{i=1}^{N} X_{p,i} + \sum_{i=1}^{N} Y_{p,i}} \tag{2}$$

We observed D^{norm} is biased by high abundance proteins and we therefore evaluate significance of a protein's sensitivity to RNase A treatment by comparing to a background of proteins with similar abundance. Specifically, we create a distribution of D^{norm} from proteins in a window surrounding protein *p* and have not been annotated as RNA-associated proteins in the literature (**equation 3**). See Figure S3A for schematic.

$$W_p = D^{norm}_{p+s}, ..., D^{norm}_{p}, ..., D^{norm}_{p-s} \quad \textbf{(3)}$$

where *s* is a window size of 100 and unannotated RNA-associated proteins are in order of abundance.

We posit that the proteins in distribution, $W_{p,}$is a mixture of unannotated RNA-associated proteins as well as non-RNA-associated proteins. In order to evaluate significance of a protein's D^{norm} being greater than what is expected by non-RNA binders, we model the distribution W_p using a two component gaussian mixture model (GMM). To ensure an accurate model fit we evaluate our GMM fit using three criteria (equation 4). First, we calculate the Baysian Information Criterion (BIC) for both the two component GMM and a one component GMM and ensure the two component GMM has a lower BIC (**equation 4a**). Second, we ensure the component with the lowest mean μ (i.e. non-RNA-associated component) has the largest weight (**equation 4b**). Finally, we ensure the largest component weight is greater than a given weight threshold t_{weight} (**equation 4c**). t_{weight} can be estimated by the expected fraction of non-RNA binders in the proteome. In practice we set t_{weight} to be between 0.6 and 0.75.

$$BIC_{2-component} \leq BIC_{1-component} \quad \textbf{(4a)}$$

$$\arg\min_{\mu} = \arg\max_{weight} \quad \textbf{(4b)}$$

$$\max_{weight} \geq t_{weight} \quad \textbf{(4c)}$$

If all three criteria are passed the lowest mean component of the two component GMM is used, otherwise the one component is used (**equation 5**).

$$GMM_p = \begin{cases} GMM_{2-component}, & \text{if criteria are met (equation 4)} \\ GMM_{1-component}, & \text{otherwise} \end{cases}$$

(5)

We next calculate the Z-score of protein *p*'s D^{norm} score relative to the non-RNA-associated component of GMM_p (**equation 6**).

$$Z_p = \frac{D_p^{norm} - \mu_{GMM_p}}{\sigma_{GMM_p}}$$

(6)

where μ_{GMM_p} and σ_{GMM_p} are the mean and standard deviation of the first component of GMM_p respectively.

Finally, we calculate a P-value of Z_p using the normal distribution survival function and then false discovery correct P-values across all proteins using the Benjamini/Hochberg correction (Benjamini and Hochberg, 1995). RNA-associated proteins were considered significant at a 0.05 FDR corrected P-value.

RNA-associated annotations, overlap comparisons and score performance analysis
Low throughput RNA binding annotations were defined as proteins with Gene Ontology (Ashburner et al., 2000) "RNA binding" annotations limited to those with evidence codes: EXP, IDA, IPI, IMP, IGI, IEP, TAS, NAS, or IC. In addition, proteins with "ribonucleoprotein" in their UniProt keywords were also included. High throughput RNA association annotations were primarily collected from Hentze *et al.* 2018 supplemental table S2 (Hentze et al., 2018). In addition, we gathered more recent high throughput datasets from (Bao et al., 2018; Huang et al., 2018; Queiroz et al., 2019; Trendel et al., 2019).

To estimate the coverage of identified RNA-associated proteins by the DIF-FRAC method independent of cell type and machine setup, the Venn diagrams in Figures 5B and S4L report only proteins with mean abundance > 10, where mean abundance is the average peptide spectral matches identified in the control and RNase A treated HEK 293T cells. To compare directly the

RNA-associated proteins identified in the high throughput sets to the DIF-FRAC method, Venn diagrams in Figure S4A-K report all proteins.

To calculate Precision vs Neg Ln P-value plots (Figure 1I and Figure S3C), we first added a pseudocount (+1e-308) to DIF-FRAC P-values and then applied -1*ln(P-value) where ln is the natural log. Precision is defined as TP/AP, where TP (true positives) is defined as proteins annotated as either high throughput or low throughput RNA binding (see above) and a Neg Ln P-value greater than a given value. AP (all predictions) is defined as any protein with a Neg Ln P-value greater than a given value. To calculate Precision vs Recall plots (Figures S3B and S3D), precision is defined above and recall is defined as TP/AKP where TP is true positives and AKP (all known positives) is defined as proteins annotated as either high throughput or low throughput RNA binding.

Classification of DIF-FRAC elution profiles

To calculate the amount a protein shifts upon RNase A treatment, we calculate the average fraction a protein is observed weighted by the PSMs observed in each fraction. The difference between the weighted average of the treated and untreated elution profiles provides the total shift amount. A protein's shift in elution from a high molecular weight to a low molecular weight results in a negative shift value whereas a shift from low molecular weight to high molecular weight corresponds to a positive value.

To calculate the amount a protein's abundance changes upon RNase A treatment, we calculate the difference of a protein's total PSMs observed in the untreated and treated samples. We further normalize this value by dividing by the sum of the total PSMs from both samples. This results in a value between 1.0 and -1.0 where a positive value corresponds to an increase in abundance upon RNase A treatment and a negative value corresponds to a decrease in abundance upon RNase A treatment.

Assembly of RNP complexes

We define the global set of RNP complexes by first creating a combined non-redundant set of CORUM (Ruepp et al., 2010) and hu.MAP (Drew et al., 2017) complexes (Jaccard coefficient < 1.0). For every complex in this global set we tested if > 50% of the protein subunits were 1) identified as an RNA-associated protein by DIF-FRAC (P-value > 0.05), 2) annotated by high throughput methods or 3) annotated by low throughput methods (see above for description of annotations). RNP Select complexes are defined as complexes whose protein subunits co-elute in the DIF-FRAC control sample (> 0.75 average Pearson correlation coefficient among subunits) and > 50% of subunits have a DIF-FRAC P-value > 0.5.

RNA-Seq Analysis

After performing quality control on the sequencing fastq files using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), 3' adapter contamination was removed using Cutadapt (v1.10) (Martin, 2011). Alignment of the 8 RNA fraction datasets was then performed with the Hisat2 transcriptome-aware aligner (v2.1.0) (Kim et al., 2015), against a Hisat2 reference index built using GRCh38/hg38 primary assembly genome fasta from Gencode (v27, Ensembl release 90) (Harrow et al., 2012) annotated with the corresponding v27 GTF (General Transfer Format) annotations. The Hisat suite Stringtie program (v1.3.3b) (Pertea et al., 2016) was used to quantify gene-level expression from the alignment files. TPM (Transcripts Per Million), a sequencing-depth-normalized estimate of reads mapping to the gene, was used for further analysis.

Data Deposition

Proteomics data and RNA-seq data will be deposited in Pride and Gene Expression Omnibus respectively upon acceptance.

Code Repository

Source code is freely available on GitHub: https://github.com/marcottelab/diffrac

References

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.-T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416–1420.

Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. *32*, D115-119.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25–29.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell *46*, 674–690.

Bao, X., Guo, X., Yin, M., Tariq, M., Lai, Y., Kanwal, S., Zhou, J., Li, N., Lv, Y., Pulido-Quetglas, C., et al. (2018). Capturing the interactome of newly transcribed RNA. Nat. Methods *15*, 213–220.

Bardoni, B., Castets, M., Huot, M.-E., Schenck, A., Adinolfi, S., Corbin, F., Pastore, A., Khandjian, E.W., and Mandel, J.-L. (2003). 82-FIP, a novel FMRP (fragile X mental retardation protein) interacting protein, shows a cell cycle-dependent intracellular localization. Hum. Mol. Genet. *12*, 1689–1698.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bleichert, F., and Baserga, S.J. (2010). Ribonucleoprotein multimers and their functions. Crit. Rev. Biochem. Mol. Biol. *45*, 331–350.

Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A.S., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwinn, M.K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. Mol. Cell *64*, 282–293.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249–253.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

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

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol. Cell *63*, 696–710.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. Mol. Cell *55*, 171–185.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinforma. Oxf. Engl. *20*, 1466–1467.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Soldà, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. *18*, 1433–1445.

Drew, K., Lee, C., Huizar, R.L., Tu, F., Borgeson, B., McWhite, C.D., Ma, Y., Wallingford, J.B., and Marcotte, E.M. (2017). Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. Mol. Syst. Biol. *13*, 932.

Escobar-Aguirre, M., Elkouby, Y.M., and Mullins, M.C. (2017). Localization in Oogenesis of Maternal Regulators of Embryonic Development. Adv. Exp. Med. Biol. *953*, 173–207.

Fang, X., Yoon, J.-G., Li, L., Tsai, Y.S., Zheng, S., Hood, L., Goodlett, D.R., Foltz, G., and Lin, B. (2011). Landscape of the SOX2 protein-protein interactome. Proteomics *11*, 921–934.

Finkelstein, J., Antony, E., Hingorani, M.M., and O'Donnell, M. (2003). Overproduction and analysis of eukaryotic multiprotein complexes in Escherichia coli using a dual-vector strategy. Anal. Biochem. *319*, 78–87.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. *45*, D190–D199.

Gerlach, P., Schuller, J.M., Bonneau, F., Basquin, J., Reichelt, P., Falk, S., and Conti, E. (2018). Distinct and evolutionary conserved structural features of the human nuclear exosome complex. ELife *7*.

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

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Hakimi, M.-A., Bochar, D.A., Schmiesing, J.A., Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K., and Shiekhattar, R. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature *418*, 994–998.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. Cell *150*, 1068–1081.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Mol. Cell *64*, 416–430.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

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

Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. Cold Spring Harb. Perspect. Med. *6*.

Holt, C.E., and Bullock, S.L. (2009). Subcellular mRNA localization in animal cells and why it matters. Science *326*, 1212–1216.

Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. Nat. Rev. Mol. Cell Biol. *7*, 529–539.

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

Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F., and Malim, M.H. (2009). RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. PLoS Pathog. *5*, e1000330.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell *162*, 425–440.

Ishikawa, K., Azuma, S., Ikawa, S., Semba, K., and Inoue, J. (2005). Identification of DRG family regulatory proteins (DFRPs): specific regulation of DRG1 and DRG2. Genes Cells Devoted Mol. Cell. Mech. *10*, 139–150.

Jung, D.-J., Sung, H.-S., Goo, Y.-W., Lee, H.M., Park, O.K., Jung, S.-Y., Lim, J., Kim, H.-J., Lee, S.-K., Kim, T.S., et al. (2002). Novel transcription coactivator complex containing activating signal cointegrator 1. Mol. Cell. Biol. *22*, 5203–5211.

Kamura, T., Maenaka, K., Kotoshiba, S., Matsumoto, M., Kohda, D., Conaway, R.C., Conaway, J.W., and Nakayama, K.I. (2004). VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. Genes Dev. *18*, 3055–3065.

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol. Cell *53*, 290–300.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem. Sci. *35*, 368–376.

Kedersha, N., Panas, M.D., Achorn, C.A., Lyons, S., Tisdale, S., Hickman, T., Thomas, M., Lieberman, J., McInerney, G.M., Ivanov, P., et al. (2016). G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. J. Cell Biol. *212*, 845–860.

Keerthivasan, G., Wickrema, A., and Crispino, J.D. (2011). Erythroblast enucleation. Stem Cells Int. *2011*, 139851.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, Y., de la Torre, A., Leal, A.A., and Finkelstein, I.J. (2017). Efficient modification of λ-DNA substrates for single-molecule studies. Sci. Rep. *7*, 2071.

Knierim, E., Hirata, H., Wolf, N.I., Morales-Gonzalez, S., Schottmann, G., Tanaka, Y., Rudnik-Schöneborn, S., Orgeur, M., Zerres, K., Vogt, S., et al. (2016). Mutations in Subunits of the Activating Signal Cointegrator 1 Complex Are Associated with Prenatal Spinal Muscular Atrophy and Congenital Bone Fractures. Am. J. Hum. Genet. *98*, 473–489.

Kobayashi, M., Figaroa, F., Meeuwenoord, N., Jansen, L.E.T., and Siegal, G. (2006). Characterization of the DNA binding and structural properties of the BRCT region of human replication factor C p140 subunit. J. Biol. Chem. *281*, 4308–4317.

Kwon, T., Choi, H., Vogel, C., Nesvizhskii, A.I., and Marcotte, E.M. (2011). MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines. J Proteome Res *10*, 2949–2958.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174–187.

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

Lingner, T., Asshauer, K.P., Schreiber, F., and Meinicke, P. (2011). CoMet--a web server for comparative functional profiling of metagenomes. Nucleic Acids Res. *39*, W518-523.

Lu, L., Lv, Y., Dong, J., Hu, S., and Peng, R. (2016). DRG1 is a potential oncogene in lung adenocarcinoma and promotes tumor progression via spindle checkpoint signaling regulation. Oncotarget *7*, 72795–72806.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.

Male, G., von Appen, A., Glatt, S., Taylor, N.M.I., Cristovao, M., Groetsch, H., Beck, M., and Müller, C.W. (2015). Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. Nat. Commun. *6*, 7387.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Matsuki, H., Takahashi, M., Higuchi, M., Makokha, G.N., Oie, M., and Fujii, M. (2013). Both G3BP1 and G3BP2 contribute to stress granule formation. Genes Cells Devoted Mol. Cell. Mech. *18*, 135–146.

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

Oron, E., Mannervik, M., Rencus, S., Harari-Steinberg, O., Neuman-Silberberg, S., Segal, D., and Chamovitz, D.A. (2002). COP9 signalosome subunits 4 and 5 regulate multiple pleiotropic pathways in Drosophila melanogaster. Dev. Camb. Engl. *129*, 4399–4409.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

Podnar, J., Deiderick, H., Huerta, G., and Hunicke-Smith, S. (2014). Next-Generation Sequencing RNA-Seq Library Construction. Curr. Protoc. Mol. Biol. *106*, 4.21.1-19.

Puente, X.S., Quesada, V., Osorio, F.G., Cabanillas, R., Cadiñanos, J., Fraile, J.M., Ordóñez, G.R., Puente, D.A., Gutiérrez-Fernández, A., Fanjul-Fernández, M., et al. (2011). Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. Am. J. Hum. Genet. *88*, 650–656.

Queiroz, R.M.L., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., et al. (2019). Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat. Biotechnol.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell *108*, 557–572.

Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.-W. (2010). CORUM: the comprehensive resource of mammalian protein complexes–2009. Nucleic Acids Res *38*, D497-501.

Sahoo, P.K., Smith, D.S., Perrone-Bizzozero, N., and Twiss, J.L. (2018). Axonal mRNA transport and translation at a glance. J. Cell Sci. *131*.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Meth *9*, 676–682.

Scotter, E.L., Chen, H.-J., and Shaw, C.E. (2015). TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. Neurother. J. Am. Soc. Exp. Neurother. *12*, 352–363.

Soll, J.M., Brickner, J.R., Mudge, M.C., and Mosammaparast, N. (2018). RNA ligase-like domain in activating signal cointegrator 1 complex subunit 1 (ASCC1) regulates ASCC complex function during alkylation damage. J. Biol. Chem. *293*, 13524–13533.

Solomon, S., Xu, Y., Wang, B., David, M.D., Schubert, P., Kennedy, D., and Schrader, J.W. (2007). Distinct structural features of caprin-1 mediate its interaction with G3BP-1 and its induction of phosphorylation of eukaryotic translation initiation factor 2alpha, entry to cytoplasmic stress granules, and selective interaction with a subset of mRNAs. Mol. Cell. Biol. *27*, 2324–2342.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb. Perspect. Biol. *3*.

Tanaka, A.J., Cho, M.T., Millan, F., Juusola, J., Retterer, K., Joshi, C., Niyazov, D., Garnica, A., Gratz, E., Deardorff, M., et al. (2015). Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. Am. J. Hum. Genet. *97*, 457–464.

Tang, Y., Wang, J., Lian, Y., Fan, C., Zhang, P., Wu, Y., Li, X., Xiong, F., Li, X., Li, G., et al. (2017). Linking long non-coding RNAs and SWI/SNF complexes to chromatin remodeling in cancer. Mol. Cancer *16*, 42.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., and Meister, G. (2017). A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. Mol. Cell *66*, 270-284.e13.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403.e19.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Vorum, H., Jacobsen, C., and Honoré, B. (2000). Calumenin interacts with serum amyloid P component. FEBS Lett. *465*, 129–134.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., et al. (2015). Panorama of ancient metazoan macromolecular complexes. Nature *525*, 339–344.

Weick, E.-M., Puno, M.R., Januszyk, K., Zinder, J.C., DiMattia, M.A., and Lima, C.D. (2018). Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. Cell *173*, 1663-1677.e21.

Weidensdorfer, D., Stöhr, N., Baude, A., Lederer, M., Köhn, M., Schierhorn, A., Buchmeier, S., Wahle, E., and Hüttelmaier, S. (2009). Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. RNA N. Y. N *15*, 104–115.

White, E.A., and Glotzer, M. (2012). Centralspindlin: at the heart of cytokinesis. Cytoskelet. Hoboken NJ *69*, 882–892.

Wood, M.P., Hollis, A., Severance, A.L., Karrick, M.L., and Schisa, J.A. (2016). RNAi Screen Identifies Novel Regulators of RNP Granules in the Caenorhabditis elegans Germ Line. G3 Bethesda Md *6*, 2643–2654.

Xie, Q., Chen, S., Tian, R., Huang, X., Deng, R., Xue, B., Qin, Y., Xu, Y., Wang, J., Guo, M., et al. (2018). Long Noncoding RNA ITPRIP-1 Positively Regulates the Innate Immune Response through Promotion of Oligomerization and Activation of MDA5. J. Virol. *92*.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851–861.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Yao, N.Y., and O'Donnell, M. (2012). The RFC clamp loader: structure and function. Subcell. Biochem. *62*, 259–279.

Yoshida, Y., Izumi, H., Torigoe, T., Ishiguchi, H., Yoshida, T., Itoh, H., and Kohno, K. (2004). Binding of RNA to p53 regulates its oligomerization and DNA-binding activity. Oncogene *23*, 4371–4379.

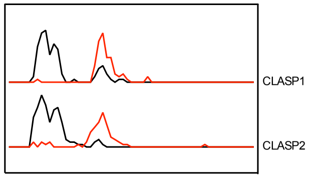
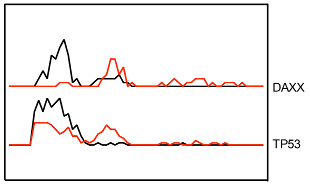
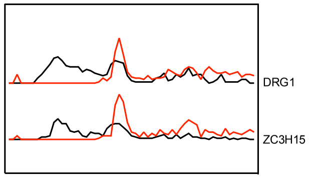
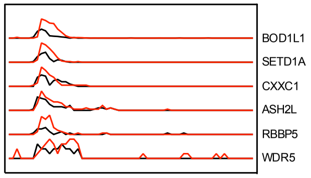
Yüce, O., Piekny, A., and Glotzer, M. (2005). An ECT2-centralspindlin complex regulates the localization and function of RhoA. J. Cell Biol. *170*, 571–582.

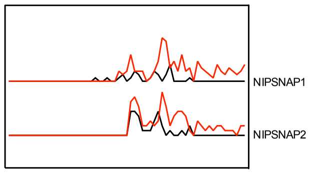
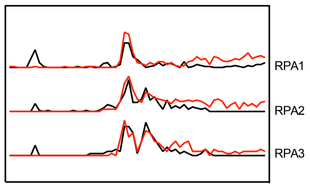
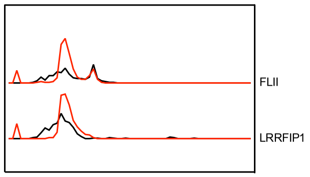
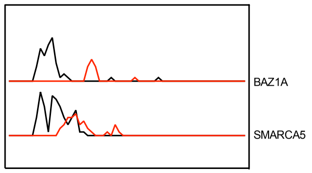
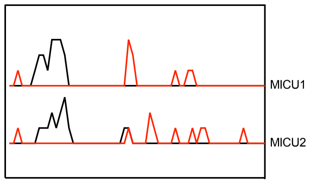
Yue, S., Serra, H.G., Zoghbi, H.Y., and Orr, H.T. (2001). The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. Hum. Mol. Genet. *10*, 25–30.

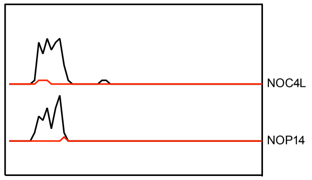
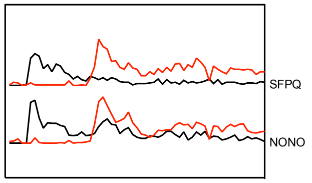
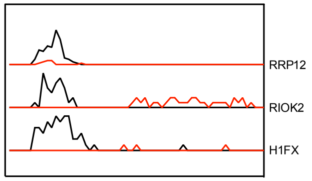
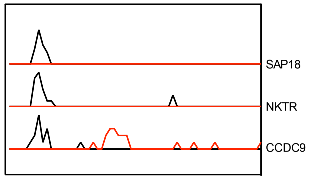
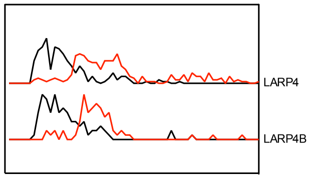
Yuzhakov, A., Kelman, Z., and O’Donnell, M. (1999). Trading places on DNA--a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. Cell *96*, 153–163.

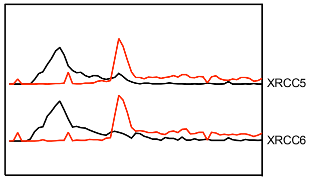
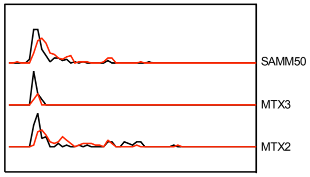
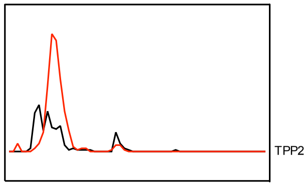
Zheng, R., Shen, Z., Tripathi, V., Xuan, Z., Freier, S.M., Bennett, C.F., Prasanth, S.G., and Prasanth, K.V. (2010). Polypurine-repeat-containing RNAs: a novel class of long non-coding RNA in mammalian cells. J. Cell Sci. *123*, 3734–3744.

Table 1: Stable RNPs identified by DIF-FRAC

Gene Names*	Complex Name	Function	Soluble without RNA? $	Disease links	CORUM /Hu.Map #	RNA.MAP id	DIF-FRAC plot	RNP class †	Refs.
CLASP1 CLASP2	N/A	Microtubule binding Microtubule dynamics	Yes	N/A	No/Yes	3807		Apo-stable	(Efimov et al., 2007)
DAXX TP53	DAXX-TP53 complex	Transcription repression	Yes	Pancreatic neuroendocrine tumors Glioblastoma multiforme Adrenocortical tumors	Yes/No	4518		Apo-stable	(Zhao et al., 2004) (Dyer et al., 2017)
DRG1 ZC3H15	Drg1/Dfrp1 complex	Microtubule binding Microtubule polymerase GTPase	Yes	Lung adenocarcinoma	No/No	4096		Apo-stable	(Lu et al., 2016) (Schellhaus et al., 2017) (Ishikawa et al., 2009)
BOD1L1 SETD1A CXXC1* ASH2L RBBP5 WDR5	SET1A/SET1B complexes	Histone methyltransferase Transcription regulation	Yes	Fanconi anaemia Mixed lineage leukemia	Yes/Yes	2005, 3307		Apo-stable	(Vedadi et al., 2017) (Higgs et al., 2015) (Brown et al., 2017)

NIPSNAP1 NIPSNAP2	N/A	Vesicular transport	No	Inflammatory pain	No/Yes	5822		Apo-stable	(Okuda-Ashitaka et al., 2012) (Yamamoto et al., 2017)
RPA1 RPA2 RPA3*	Replication protein A complex	Single-stranded DNA binding DNA metabolism	Yes	Werner Syndrome	Yes/Yes	3204		Apo-stable	(Machwe et al., 2011) (Fan and Pavletich, 2012)
FLII LRRFIP1	FLII-LRRFIP1 complex	Transcriptional activation Actin binding	Yes	Prostate cancer	No/Yes	3626		Apo-stable	(Wilson et al., 1998) (Wang et al., 2016)
BAZ1A SMARCA5	WCRF complex	Chromatin remodeling	No	Intellectual disability	Yes/No	2105		Compositional	(Bochar et al., 2000) (Zaghlool et al., 2016)
MICU1* MICU2*	MICU1-MICU2 heterodimer	Calcium ion transport	Yes	Myopathy with extrapyramidal signs	Yes/Yes	4318		Structural	(Logan et al., 2014)

NOC4L NOP14	N/A	Ribosome processing and biogenesis	No	Recurrent pregnancy loss	No/Yes	4220		Structural	(Suzuki et al., 2018)
SFPQ NONO	PSF-p54(nrb) complex	Spicing factor DNA recombination	Yes	Intellectual disability	Yes/Yes	327		Apo-stable	(Bladen et al., 2005) (Mircsof et al., 2015)
RRP12 RIOK2 H1FX	N/A	rRNA processing	No	N/A	No/Yes	5795		Structural	(Zemp et al., 2009)
SAP18 NKTR CCDC9	N/A	SAP18 is involved in RNA processing and splicing NKTR is involved in protein peptidyl-prolyl isomerization	No	N/A	No/Yes	6027		Structural	(Davis et al., 2010)
LARP4 LARP4B	N/A	Translation regulation	Yes	N/A	No/Yes	3327		Apo-stable	(Yang et al., 2011) (Schaffler et al., 2010)

XRCC5 XRCC6	Ku antigen complex	DNA damage and repair	Yes	Systemic lupus erythematosus	Yes/No	2930		Apo-stable	(Spagnolo et al., 2006)
SAMM50* MTX3* MTX2	N/A	Protein transport	Yes	N/A	No/Yes	1450		Apo-stable	
TPP2	Tripeptidyl-peptidase II	Serine protease	Yes	Muscle wasting Obesity Cancer	Yes/No			Apo-stable	(Schonegge et al., 2012) (Rockel et al., 2012)

* RNA-associated proteins newly identified by DIF-FRAC are shown in red. Those with an asterisk are above the 5% FDR cutoff. RNA-associated proteins in black have been previously annotated or identified by high-throughput methods (See Table S1).

# Evidence for all or some protein complex subunits interacting in CORUM or Hu.Map

$ Insolubility in the absence of RNA is inferred by an increase in apparent molecular weight of the complex upon RNA digestion, or a complete disappearance of signal. This is consistent with the RNP being solubilized by RNA, as suggested by Maharana et al (Maharana et al., 2018)

† RNP classes apo-stable, structural, or compositional as described in Fig. 4.

Fig1

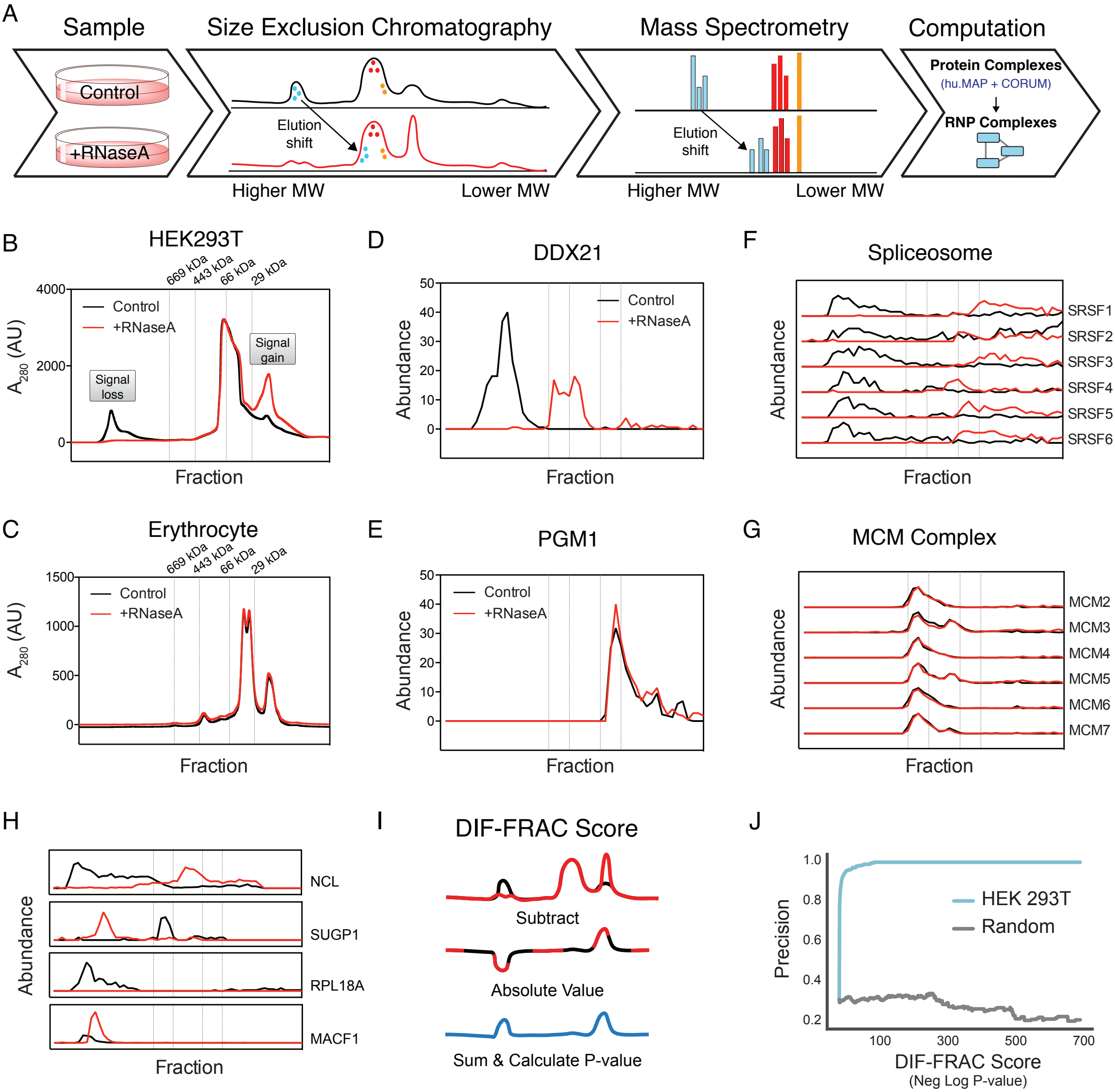

A
Sample
Size Exclusion Chromatography
Mass Spectrometry
Computation
Control
+RNaseA
Elution shift
Higher MW
Lower MW
Elution shift
Higher MW
Lower MW
Protein Complexes
(hu.MAP + CORUM)
RNP Complexes
B
HEK293T
669 kDa
443 kDa
66 kDa
29 kDa
4000
2000
0
A280 (AU)
Control
+RNaseA
Signal loss
Signal gain
Fraction
C
Erythrocyte
669 kDa
443 kDa
66 kDa
29 kDa
1500
1000
500
0
A280 (AU)
Control
+RNaseA
Fraction
D
DDX21
50
40
30
20
10
0
Abundance
Control
+RNaseA
Fraction
E
PGM1
50
40
30
20
10
0
Abundance
Control
+RNaseA
Fraction
F
Spliceosome
Abundance
SRSF1
SRSF2
SRSF3
SRSF4
SRSF5
SRSF6
Fraction
G
MCM Complex
Abundance
MCM2
MCM3
MCM4
MCM5
MCM6
MCM7
Fraction
H
Abundance
NCL
SUGP1
RPL18A
MACF1
Fraction
I
DIF-FRAC Score
Subtract
Absolute Value
Sum & Calculate P-value
J
Precision
1.0
0.8
0.6
0.4
0.2
HEK 293T
Random
100
300
500
700
DIF-FRAC Score
(Neg Log P-value)

Fig2

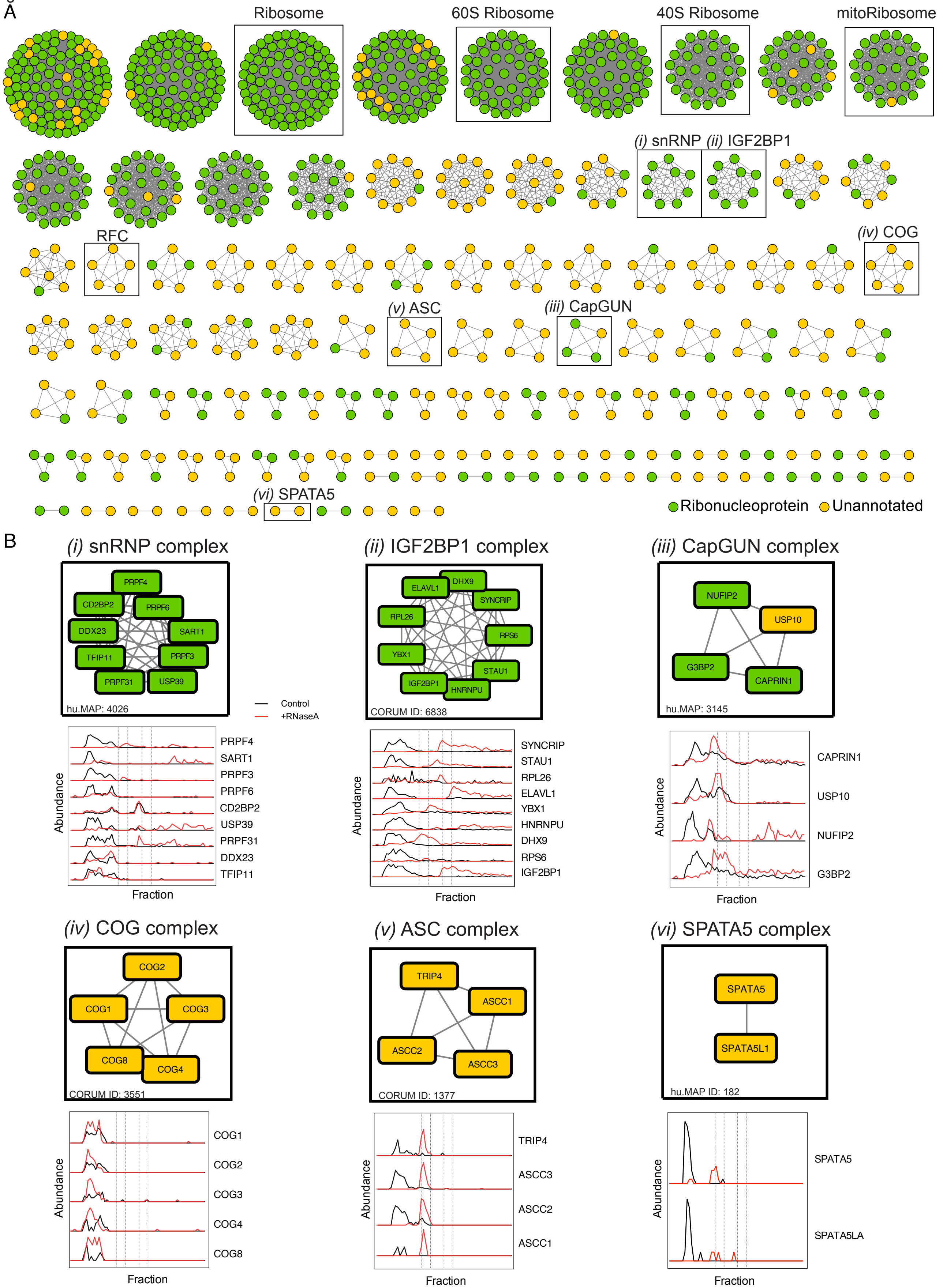

A
Ribosome
60S Ribosome
40S Ribosome
mitoRibosome
(i) snRNP
(ii) IGF2BP1
RFC
(iv) COG
(v) ASC
(iii) CapGUN
(vi) SPATA5
● Ribonucleoprotein
● Unannotated
B
(i) snRNP complex
PRPF4
CD2BP2
PRPF6
DDX23
SART1
TFIP11
PRPF3
PRPF31
USP39
hu.MAP: 4026
Control
+RNaseA
Abundance
PRPF4
SART1
PRPF3
PRPF6
CD2BP2
USP39
PRPF31
DDX23
TFIP11
Fraction
(ii) IGF2BP1 complex
ELAVL1
DHX9
SYNCRIP
RPL26
RPS6
YBX1
STAU1
IGF2BP1
HNRNPU
CORUM ID: 6838
Abundance
SYNCRIP
STAU1
RPL26
ELAVL1
YBX1
HNRNPU
DHX9
RPS6
IGF2BP1
Fraction
(iii) CapGUN complex
NUFIP2
USP10
G3BP2
CAPRIN1
hu.MAP: 3145
Abundance
CAPRIN1
USP10
NUFIP2
G3BP2
Fraction
(iv) COG complex
COG2
COG1
COG3
COG8
COG4
CORUM ID: 3551
Abundance
COG1
COG2
COG3
COG4
COG8
Fraction
(v) ASC complex
TRIP4
ASCC1
ASCC2
ASCC3
CORUM ID: 1377
Abundance
TRIP4
ASCC3
ASCC2
ASCC1
Fraction
(vi) SPATA5 complex
SPATA5
SPATA5L1
hu.MAP ID: 182
Abundance
SPATA5
SPATA5LA
Fraction

Fig3

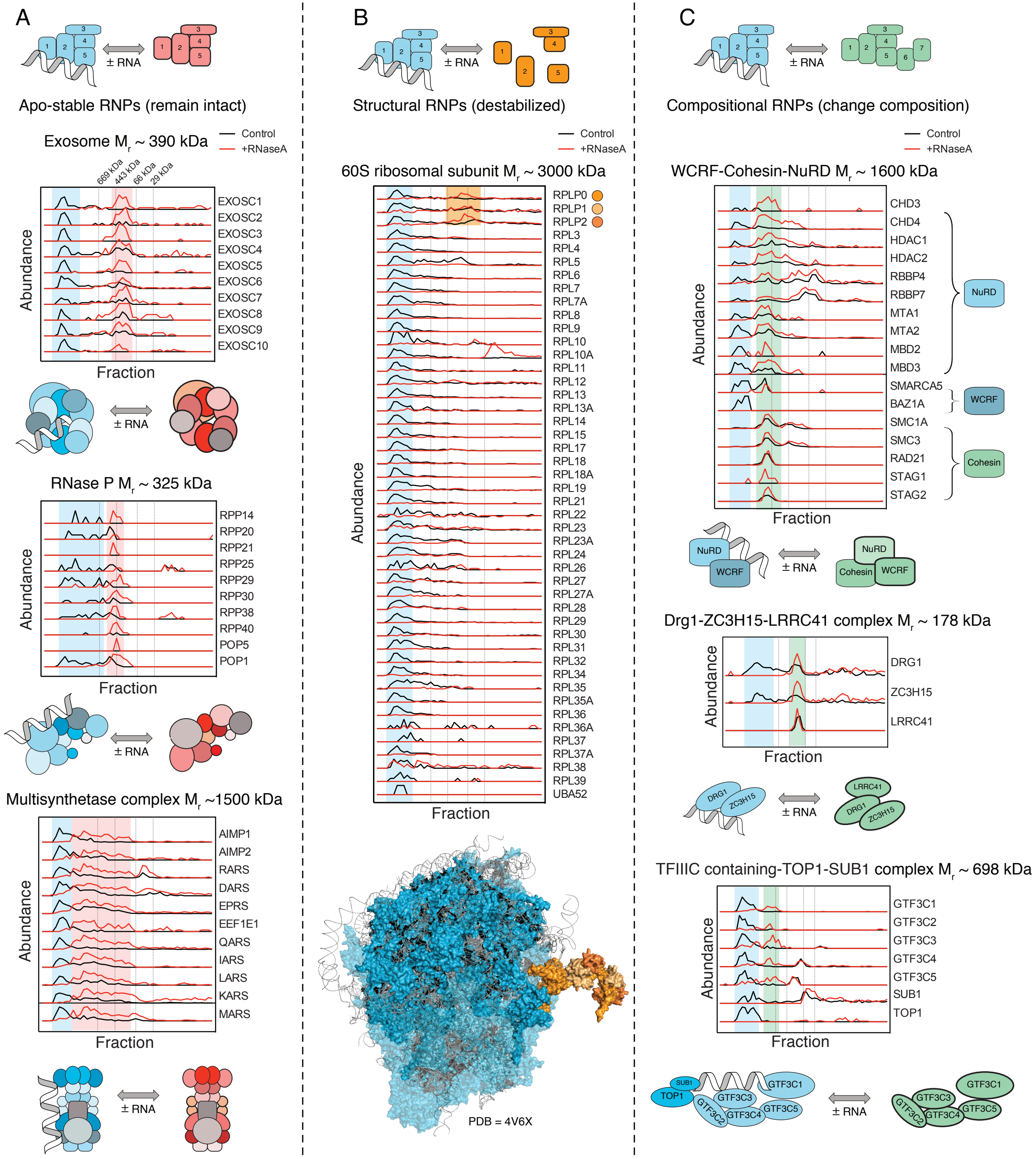

A
1
2
3
4
5
± RNA
Apo-stable RNPs (remain intact)
Exosome M$_r$ ~ 390 kDa
Control
+RNaseA
669 kDa
443 kDa
66 kDa
29 kDa
Abundance
EXOSC1
EXOSC2
EXOSC3
EXOSC4
EXOSC5
EXOSC6
EXOSC7
EXOSC8
EXOSC9
EXOSC10
Fraction
± RNA
RNase P M$_r$ ~ 325 kDa
Abundance
RPP14
RPP20
RPP21
RPP25
RPP29
RPP30
RPP38
RPP40
POP5
POP1
Fraction
± RNA
Multisynthetase complex M$_r$ ~1500 kDa
Abundance
AIMP1
AIMP2
RARS
DARS
EPRS
EEF1E1
QARS
IARS
LARS
KARS
MARS
Fraction
± RNA
B
1
2
3
4
5
± RNA
Structural RNPs (destabilized)
Control
+RNaseA
60S ribosomal subunit M$_r$ ~ 3000 kDa
Abundance
RPLP0
RPLP1
RPLP2
RPL3
RPL4
RPL5
RPL6
RPL7
RPL7A
RPL8
RPL9
RPL10
RPL10A
RPL11
RPL12
RPL13
RPL13A
RPL14
RPL15
RPL17
RPL18
RPL18A
RPL19
RPL21
RPL22
RPL23
RPL23A
RPL24
RPL26
RPL27
RPL27A
RPL28
RPL29
RPL30
RPL31
RPL32
RPL34
RPL35
RPL35A
RPL36
RPL36A
RPL37
RPL37A
RPL38
RPL39
UBA52
Fraction
PDB = 4V6X
C
1
2
3
4
5
6
7
± RNA
Compositional RNPs (change composition)
Control
+RNaseA
WCRF-Cohesin-NuRD M$_r$ ~ 1600 kDa
Abundance
CHD3
CHD4
HDAC1
HDAC2
RBBP4
RBBP7
MTA1
MTA2
MBD2
MBD3
SMARCA5
BAZ1A
SMC1A
SMC3
RAD21
STAG1
STAG2
NuRD
WCRF
Cohesin
Fraction
NuRD
WCRF
± RNA
NuRD
Cohesin
WCRF
Drg1-ZC3H15-LRRC41 complex M$_r$ ~ 178 kDa
Abundance
DRG1
ZC3H15
LRRC41
Fraction
DRG1
ZC3H15
± RNA
LRRC41
DRG1
ZC3H15
TFIIIC containing-TOP1-SUB1 complex M$_r$ ~ 698 kDa
Abundance
GTF3C1
GTF3C2
GTF3C3
GTF3C4
GTF3C5
SUB1
TOP1
Fraction
SUB1
TOP1
GTF3C1
GTF3C3
GTF3C2
GTF3C4
GTF3C5
± RNA
GTF3C1
GTF3C3
GTF3C2
GTF3C4
GTF3C5

Fig4

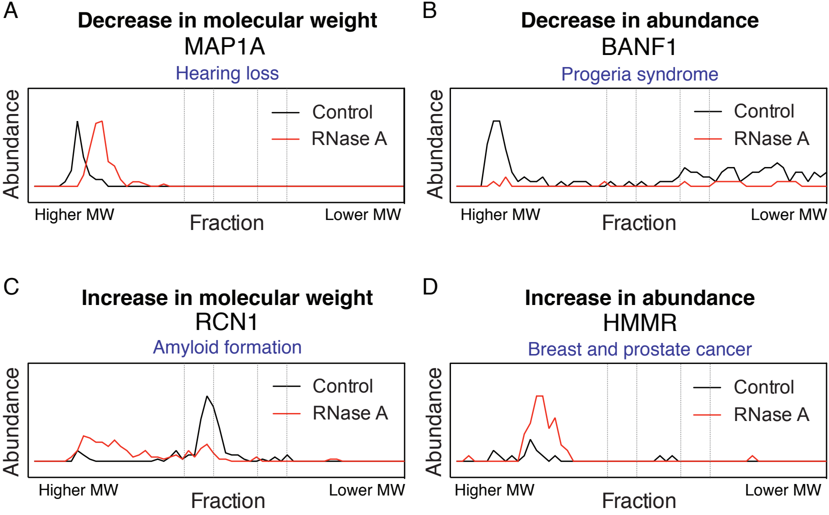

A
Decrease in molecular weight
MAP1A
Hearing loss
Abundance
Control
RNase A
Higher MW
Fraction
Lower MW
B
Decrease in abundance
BANF1
Progeria syndrome
Abundance
Control
RNase A
Higher MW
Fraction
Lower MW
C
Increase in molecular weight
RCN1
Amyloid formation
Abundance
Control
RNase A
Higher MW
Fraction
Lower MW
D
Increase in abundance
HMMR
Breast and prostate cancer
Abundance
Control
RNase A
Higher MW
Fraction
Lower MW

Fig5

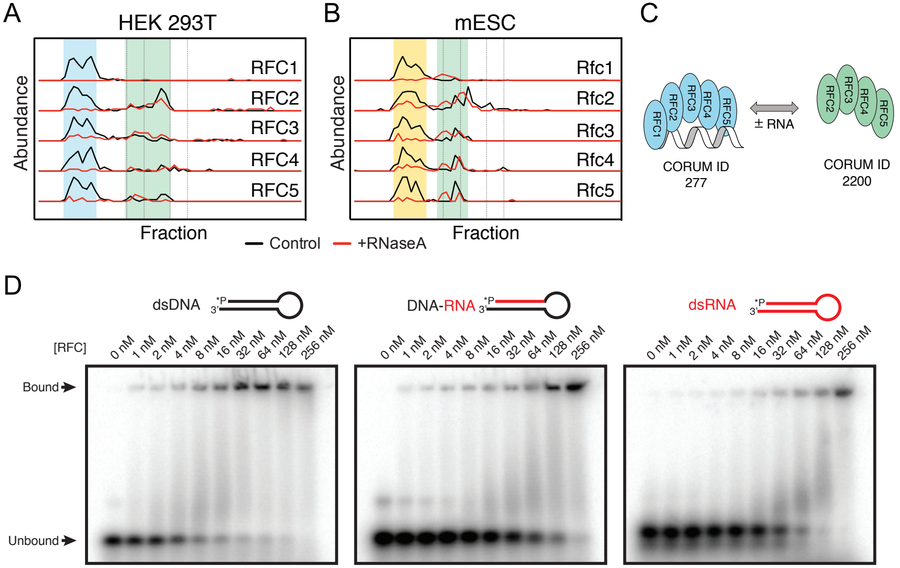
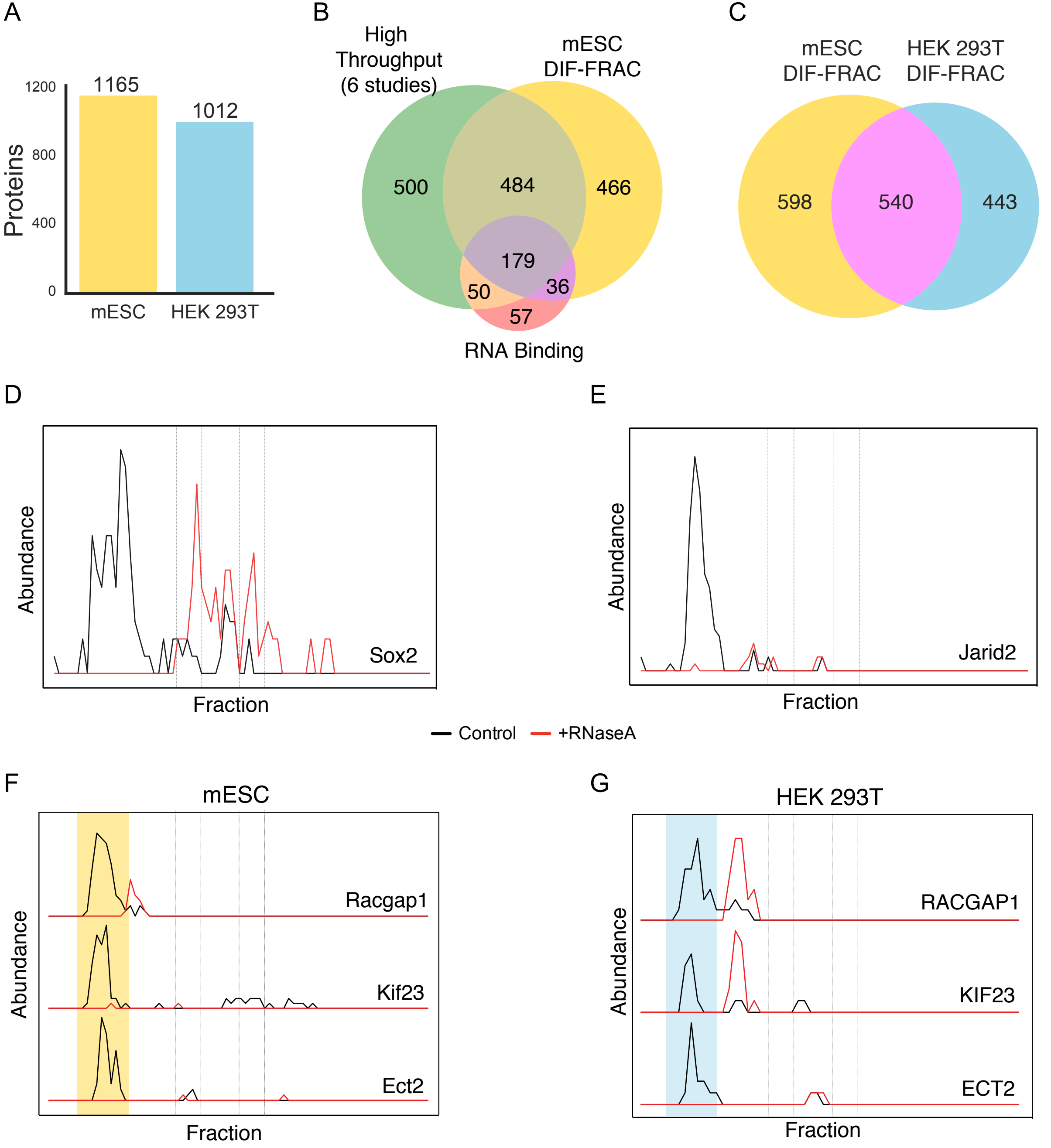
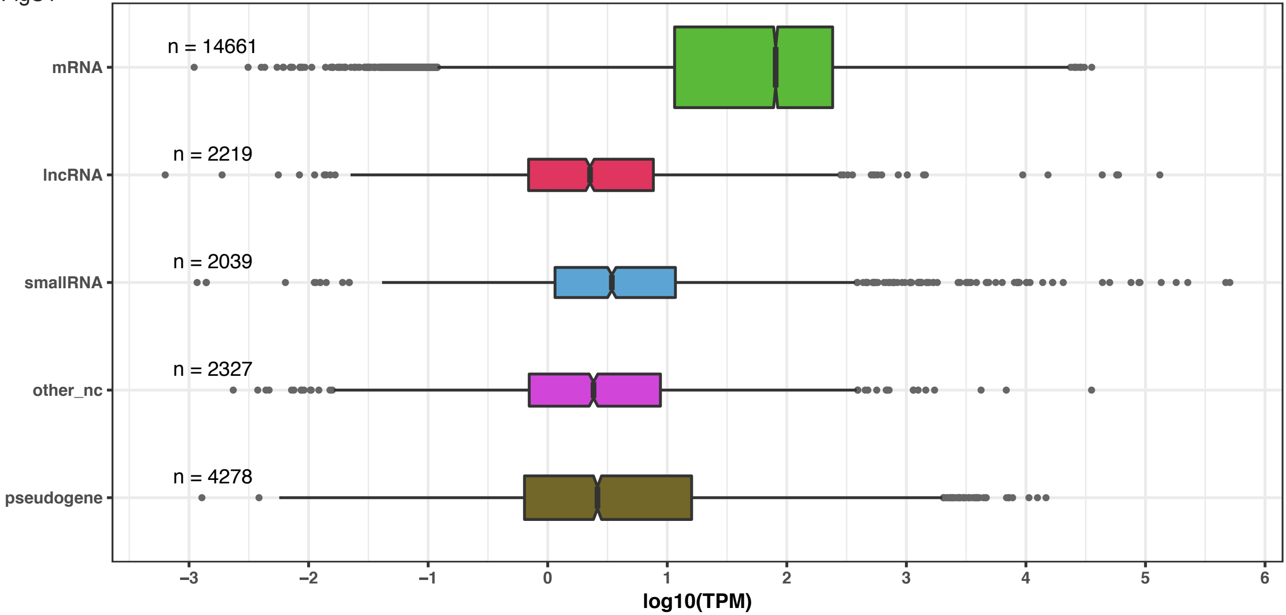

A
HEK 293T
RFC1
RFC2
RFC3
RFC4
RFC5
Abundance
Fraction
B
mESC
Rfc1
Rfc2
Rfc3
Rfc4
Rfc5
Abundance
Fraction
Control
+RNaseA
C
RFC1
RFC2
RFC3
RFC4
RFC5
CORUM ID
277
± RNA
RFC2
RFC3
RFC4
RFC5
CORUM ID
2200
D
dsDNA
3' *P
DNA-RNA
3' *P
dsRNA
3' *P
[RFC]
0 nM
1 nM
2 nM
4 nM
8 nM
16 nM
32 nM
64 nM
128 nM
256 nM
Bound
Unbound

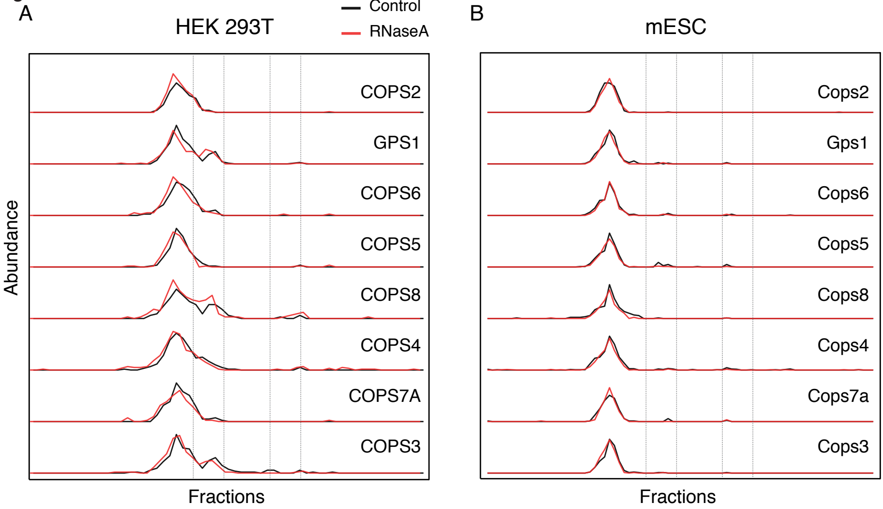
Fig6


A
Proteins
1200
800
400
0
1165
1012
mESC
HEK 293T
B
High Throughput (6 studies)
mESC DIF-FRAC
RNA Binding
500
484
466
179
50
36
57
C
mESC DIF-FRAC
HEK 293T DIF-FRAC
598
540
443
D
Abundance
Fraction
Sox2
E
Abundance
Fraction
Jarid2
Control
+RNaseA
F
mESC
Abundance
Fraction
Racgap1
Kif23
Ect2
G
HEK 293T
Abundance
Fraction
RACGAP1
KIF23
ECT2

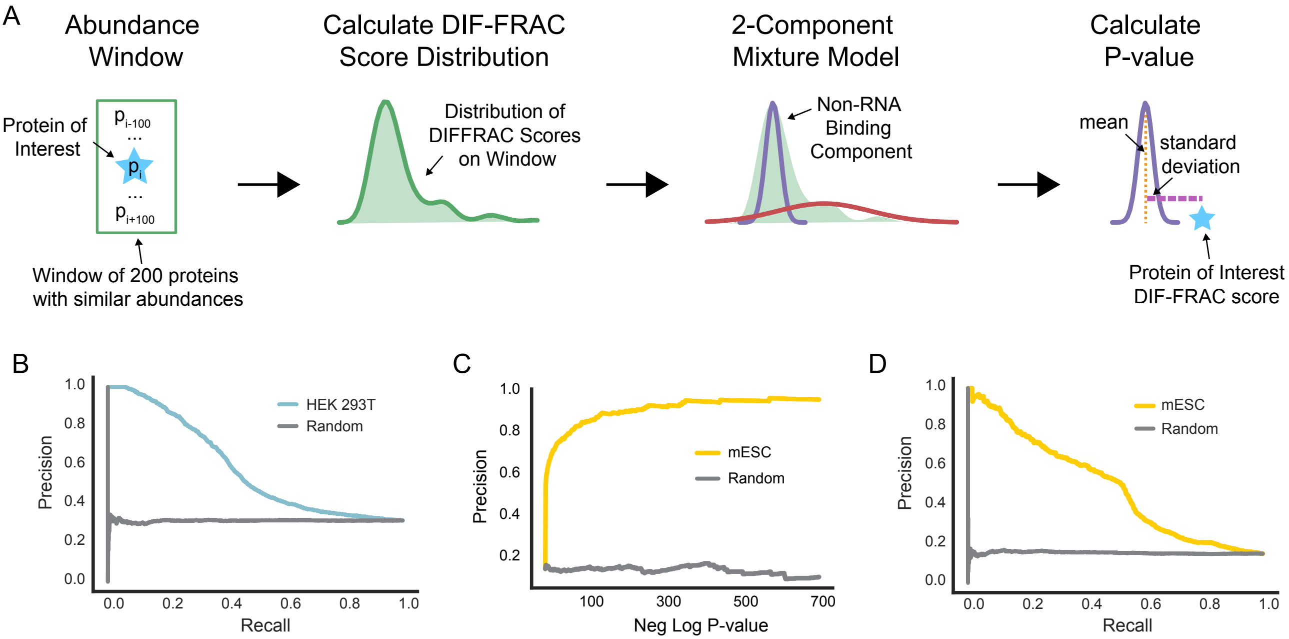
FigS1


n = 14661
n = 2219
n = 2039
n = 2327
n = 4278
mRNA
lncRNA
smallRNA
other_nc
pseudogene
−3
−2
−1
0
1
2
3
4
5
6
log10(TPM)

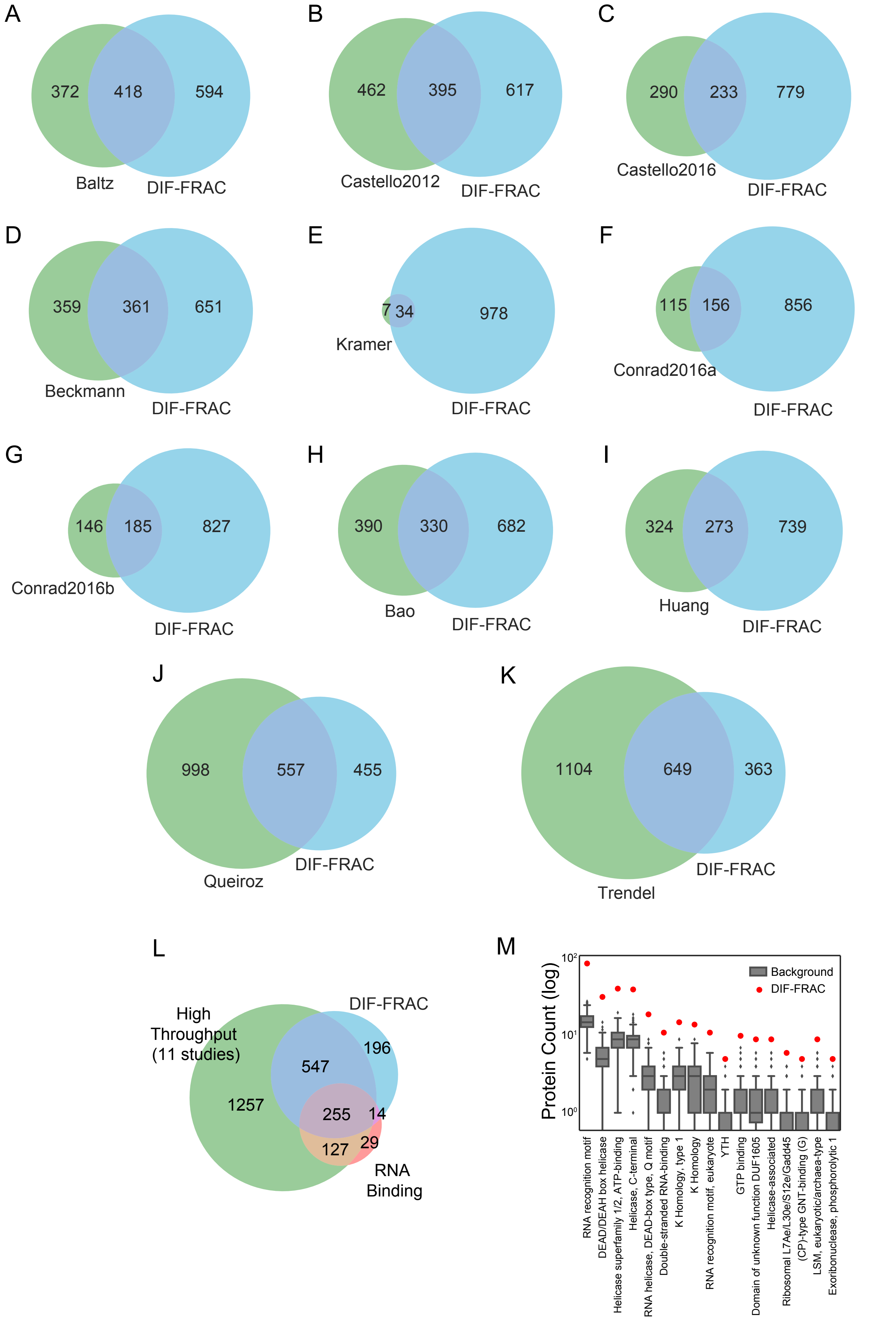
FigS2


A
HEK 293T
Control
RNaseA
Abundance
COPS2
GPS1
COPS6
COPS5
COPS8
COPS4
COPS7A
COPS3
Fractions
B
mESC
Cops2
Gps1
Cops6
Cops5
Cops8
Cops4
Cops7a
Cops3
Fractions

FigS3


A
Abundance Window
Protein of Interest
p_{i-100}
...
p_i
...
p_{i+100}
Window of 200 proteins with similar abundances
Calculate DIF-FRAC Score Distribution
Distribution of DIFFRAC Scores on Window
2-Component Mixture Model
Non-RNA Binding Component
Calculate P-value
mean
standard deviation
Protein of Interest DIF-FRAC score
B
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
HEK 293T
Random
C
Precision
Neg Log P-value
1.0
0.8
0.6
0.4
0.2
100
300
500
700
mESC
Random
D
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
mESC
Random

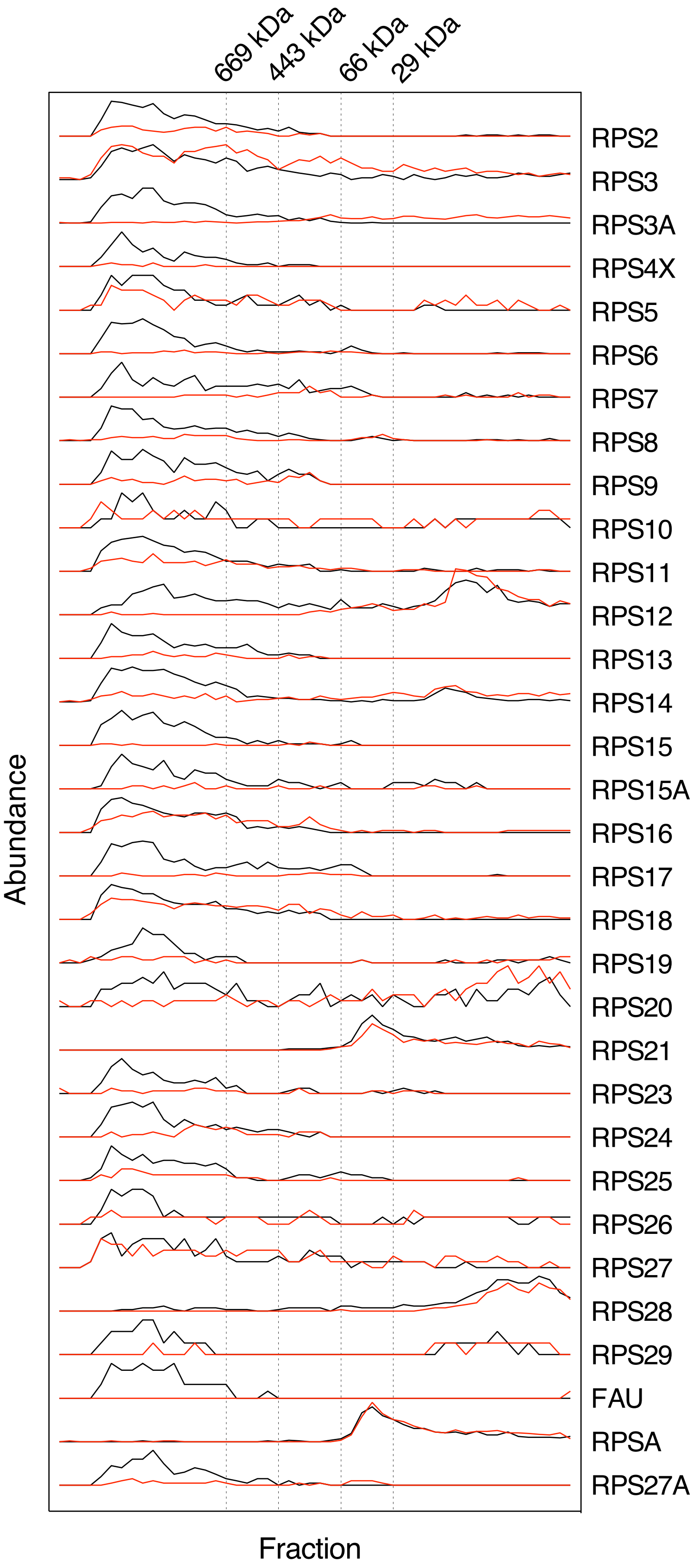
FigS4



A
372
418
594
Baltz
DIF-FRAC
B
462
395
617
Castello2012
DIF-FRAC
C
290
233
779
Castello2016
DIF-FRAC
D
359
361
651
Beckmann
DIF-FRAC
E
7
34
978
Kramer
DIF-FRAC
F
115
156
856
Conrad2016a
DIF-FRAC
G
146
185
827
Conrad2016b
DIF-FRAC
H
390
330
682
Bao
DIF-FRAC
I
324
273
739
Huang
DIF-FRAC
J
998
557
455
Queiroz
DIF-FRAC
K
1104
649
363
Trendel
DIF-FRAC
L
High Throughput (11 studies)
DIF-FRAC
1257
547
196
255
14
127
29
RNA Binding
M
Protein Count (log)
10^2
10^1
10^0
Background
DIF-FRAC
RNA recognition motif
DEAD/DEAH box helicase
Helicase superfamily 1/2, ATP-binding
Helicase, C-terminal
RNA helicase, DEAD-box type, Q motif
Double-stranded RNA-binding
K Homology, type 1
K Homology
RNA recognition motif, eukaryote
YTH
GTP binding
Domain of unknown function DUF1605
Helicase-associated
Ribosomal L7Ae/L30e/S12e/Gadd45
(CP)-type GNT-binding (G)
LSM, eukaryotic/archaea-type
Exoribonuclease, phosphorolytic 1

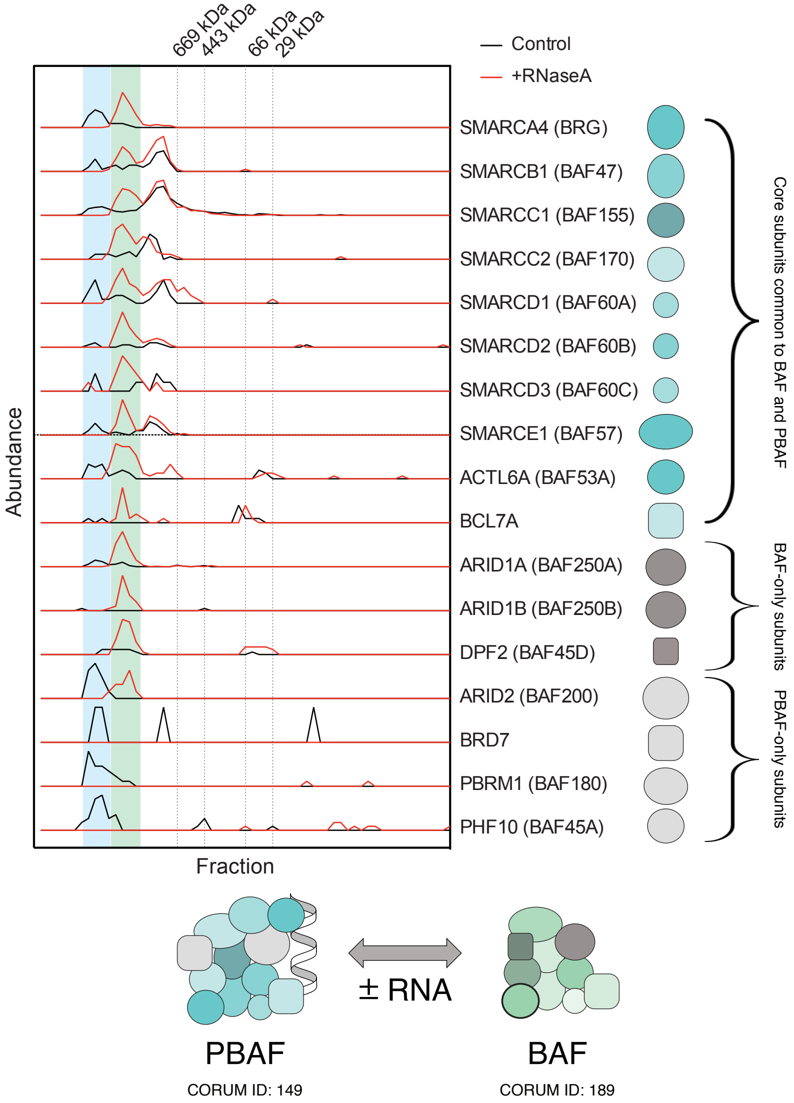
FigS5

40S ribosomal subunit (~1500 kDa)

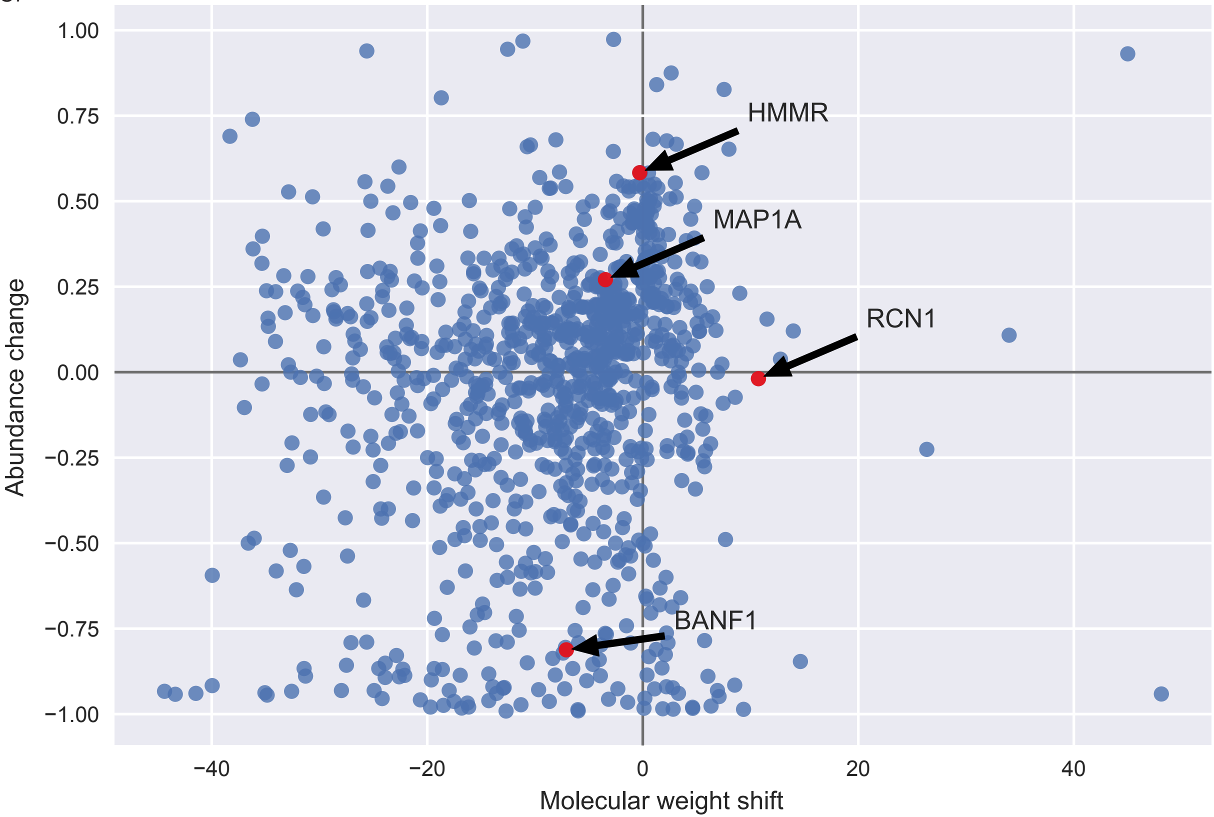


FigS6

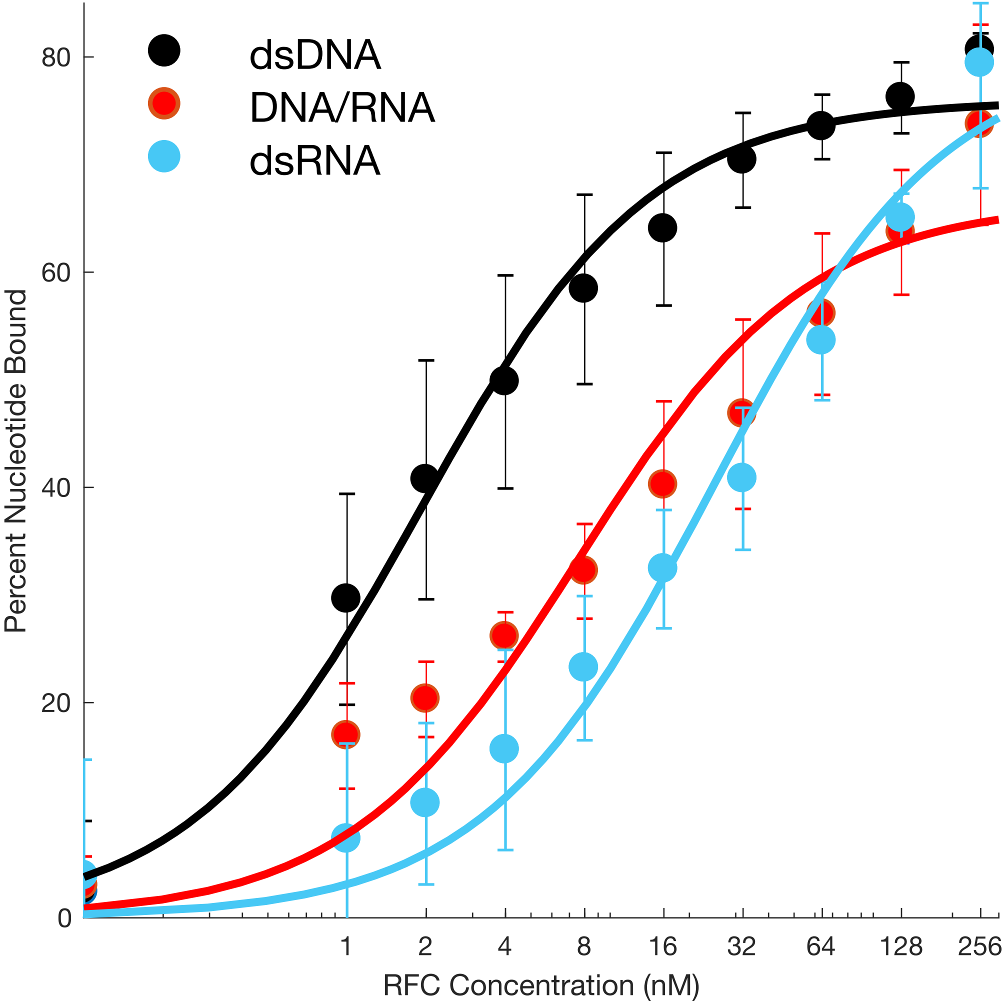
BAF and PBAF complexes M ~ 2000 kDa



FigS7


HMMR
MAP1A
RCN1
BANF1
Abundance change
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Molecular weight shift
−40
−20
0
20
40

FigS8


dsDNA
DNA/RNA
dsRNA
Percent Nucleotide Bound
0
20
40
60
80
1
2
4
8
16
32
64
128
256
RFC Concentration (nM)