A systematic, label-free method for identifying RNA-associated proteins *in vivo* provides insights into vertebrate ciliary beating machinery

Kevin Drew [1], Chanjae Lee [1], Rachael M. Cox, Vy Dang, Caitlin C. Devitt, Claire D. McWhite, Ophelia Papoulas, Ryan L. Huizar, Edward M. Marcotte [**], John B. Wallingford [*]

Dept. of Molecular Biosciences, Center for Systems and Synthetic Biology, University of Texas, Austin, TX, 78712, USA

A B S T R A C T

Cell-type specific RNA-associated proteins are essential for development and homeostasis in animals. Despite a massive recent effort to systematically identify RNA-associated proteins, we currently have few comprehensive rosters of cell-type specific RNA-associated proteins in vertebrate tissues. Here, we demonstrate the feasibility of determining the RNA-associated proteome of a defined vertebrate embryonic tissue using DIF-FRAC, a systematic and universal (*i.e.*, label-free) method. Application of DIF-FRAC to cultured tissue explants of *Xenopus* mucociliary epithelium identified dozens of known RNA-associated proteins as expected, but also several novel RNA-associated proteins, including proteins related to assembly of the mitotic spindle and regulation of ciliary beating. In particular, we show that the inner dynein arm tether Cfap44 is an RNA-associated protein that localizes not only to axonemes, but also to liquid-like organelles in the cytoplasm called DynAPs. This result led us to discover that DynAPs are generally enriched for RNA. Together, these data provide a useful resource for a deeper understanding of mucociliary epithelia and demonstrate that DIF-FRAC will be broadly applicable for systematic identification of RNA-associated proteins from embryonic tissues.

1. Introduction

RNA-associated proteins play diverse roles in normal cellular physiology and their disruption is linked to diverse pathologies (Castello et al., 2013; Gerstberger et al., 2014; Hentze et al., 2018; Quinn and Chang, 2016; Wickramasinghe and Venkitaraman, 2016). Encompassing both direct RNA-binding proteins and indirect interactors, the RNA-associated protein universe is expanding rapidly. Indeed, there is now a global effort to systematically identify and characterize RNA-associated proteins, and this effort continues to provide new biological insights as well as invaluable resources for hypothesis generation (e.g. (Baltz et al., 2012; Bao et al., 2018; Brannan et al., 2016; Castello et al., 2016; Caudron-Herger et al., 2019; He et al., 2016; Huang et al., 2018; Mallam et al., 2019; Queiroz et al., 2019; Treiber et al., 2017; Trendel et al., 2019)).

However, while this effort has been focused largely on a narrow selection of cultured cell types, a wide array of *cell-type specific* RNA-associated proteins are known to be essential for normal development and homeostasis. For example, RNA-associated proteins control the specific transport and localization of mRNAs in cells ranging from totipotent fertilized eggs to highly differentiated cells such as neurons (Holt and Bullock, 2009; Medioni et al., 2012; Sahoo et al., 2018). Moreover, mammalian ribosomal proteins are now thought to play cell-type specific roles, for example in the translation of developmental regulators such as Hox genes during mouse development (Kondrashov et al., 2011; Xue and Barna, 2012), in development of the spleen (Bolze et al., 2013), and in hematopoiesis (Choesmel et al., 2007). Recent work in *C. elegans* has argued against specialized ribosomes (Cenik et al., 2019; Haag and Dinman, 2019), but *C. elegans* nonetheless displays a curious non-cell-autonomous requirement for ribosomes in growth control (Artiles et al., 2019; Cenik et al., 2019).

At the same time, RNA-associated proteins have also emerged as critical factors for the assembly of liquid-like organelles, including not only ubiquitous organelles such as nucleoli and stress granules, but also cell-type specific organelles such as Cajal bodies (Banani et al., 2017; Lin et al., 2015; Mittag and Parker, 2018; Sawyer et al., 2016). In a recent

---

study, we described novel liquid-like organelles called Dyn-APs that are specific to multiciliated cells, where they control the assembly of dynein motors that drive ciliary beating (Huizar et al., 2018). DynAPs are present in motile ciliated cells of *Xenopus*, zebrafish, and mammals, and their disruption is associated with human motile ciliopathy (Horani et al., 2018; Huizar et al., 2018; Li et al., 2017; Mali et al., 2018). Thus, understanding the role of RNA-associated proteins in DynAPs is an important specific challenge, while development of methods for systematic identification of RNA-associated proteins from animal tissues will be broadly useful.

Here, we demonstrate the feasibility of determining the RNA-associated proteome of a defined vertebrate embryonic tissue using DIF-FRAC, a systematic and universal (*i.e.,* label-free) method (Mallam et al., 2019). Application of DIF-FRAC to cultured tissue explants of *Xenopus* mucociliary epithelium identified over 380 RNA-associated proteins, including dozens of known RNA-associated proteins, but also several novel RNA-associated proteins. Many of the novel RNA-associated proteins

play roles in assembly of the mitotic spindle and the regulation of ciliary beating. These data provide a useful resource for a deeper understanding of mucociliary epithelia and demonstrate that DIF-FRAC will be broadly applicable for systematic identification of RNA-associated proteins in both model and non-model organisms.

2. Results

We recently developed a novel method for systematically identifying RNA-protein interactions based on differential fractionation (DIF-FRAC) (Fig. 1) (Mallam et al., 2019). In DIF-FRAC, a protein lysate from theoretically any native or non-denatured sample is split into two; one is treated with RNAse and the other serves as a control; each sample is then independently subjected to size-exclusion chromatography (SEC) and the contents of each fraction are independently quantified by mass spectrometry. Shifts in elution profiles reveal alterations to protein complexes in the presence or absence of RNA, and a novel statistical framework is
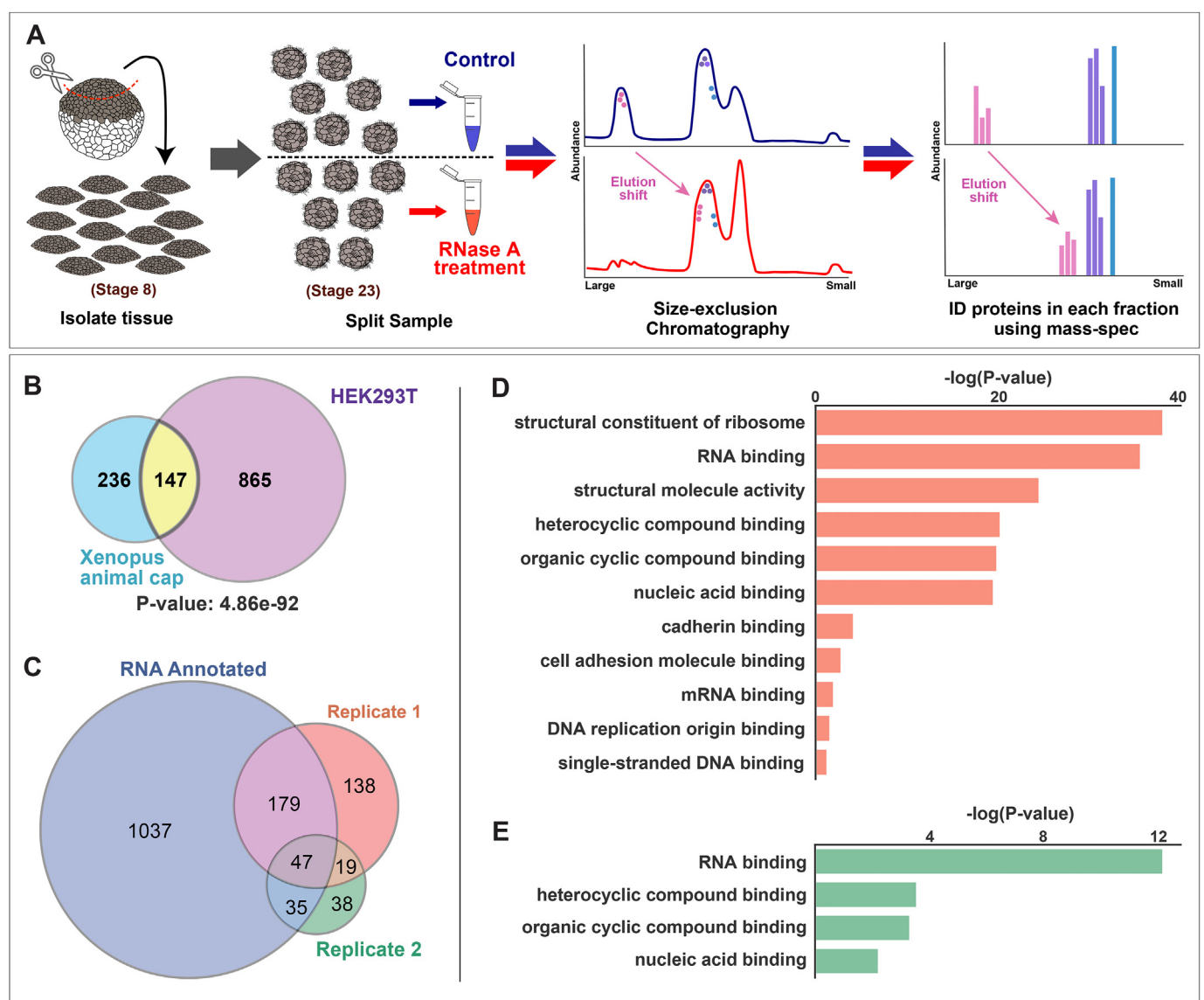

Fig. 1. **Differential Fractionation (DIF-FRAC) identifies RNA-associated proteins in a mucociliary epithelium.** (A) Experimental workflow of the RNAse DIF-FRAC experiment on *Xenopus* animal cap explants. (B) Venn diagram displaying overlap of RNA-associated proteins in *Xenopus* animal caps with previously published data from HEK293T cells (Mallam et al., 2019). The *p*-value represents the probability of overlap based on chance using the hypergeometric test. (C) Venn diagram of high confidence hits identified in replicate experiments and previously annotated RNA associated proteins. The set of previously annotated RNA associated proteins was constructed by including those with >10 peptide spectral matches in either of the replicates. (D–E) Gene ontology molecular function enrichment analysis of high confidence DIF-FRAC hits from replicate 1 (D) and replicate 2 (E).

used to quantify these changes and thereby identify RNA-associated proteins (Fig. 1A).

To test the efficacy of DIF-FRAC on embryonic tissue, we used the pluripotent ectoderm from blastula stage *Xenopus* embryos (so-called "animal caps"), which can be readily transformed into a wide array of organs and tissues, including a motile ciliated epithelium (Ariizumi et al., 2009; Walentek and Quigley, 2017; Werner and Mitchell, 2012). This tissue can be obtained in abundance, as we and others have demonstrated in large-scale genomic studies of ciliogenesis and cilia function (Chung et al., 2014; Ma et al., 2014; Quigley and Kintner, 2017). As outlined in Fig. 1A, we collected ~3000 micro-dissected animal caps and cultured them to stage 23, when motile cilia have been assembled and begin to beat. We performed DIF-FRAC using ~80 SEC fractions each for control and RNase treated samples.

Due to the allotetraploid nature of the *Xenopus laevis* genome (Session et al., 2016), multiple copies of most genes exist in the genome, hindering accurate identification of proteins by mass spectrometry. Previous approaches used a transcriptome-based reference proteome (Wühr et al., 2014), but this approach is limited to proteins whose transcripts were identified in the employed mRNA sequencing datasets and therefore under-represent proteins with highly restricted expression, for example those related to ciliary motility.

To overcome these limitations, we applied an orthology-based approach that we recently developed for proteomic comparisons across polyploid plant species (McWhite et al., 2020), in which we collapse highly related protein sequences into eggNOG vertebrate orthology groups (Huerta-Cepas et al., 2016) (Supp. Fig. 1, and see methods). Using this approach, we substantially increased the number of uniquely assigned mass spectra, as well as the number of identified orthology groups compared to the standard *Xenopus laevis* reference proteome (Supp. Table 1). Peptide spectral matches (PSMs) for each orthology group were then used to construct elution profiles (*i.e.*, abundance of protein across fractions) and elution profiles were compared across control and RNase treatments.

To identify RNA-associated proteins in DIF-FRAC data, we use a computational framework to identify high confidence changes in a protein's elution behavior (the "DIF-FRAC score") which allows us to assign *Z-scores* by comparing each protein's DIF-FRAC score to an abundance-controlled background distribution of DIF-FRAC scores from non-RNA-associated proteins. (A fuller explanation can be found in the Methods section and in our published work (Mallam et al., 2019).) This *Z-score* takes into account the observed abundance as well as the size of the elution profile shift. This statistical framework provides a ranked confidence score for each protein.

Using this approach, we identified over 380 RNA-associated proteins in *Xenopus* animal caps, and this set markedly overlapped the set previously identified in HEK293T cells (Mallam et al., 2019) (Fig. 1B) (Supp. Table 2). To determine the robustness of the DIF-FRAC method, we analyzed an independent biological replicate of our experiment. An examination of the correlation of protein abundances (PSM) across replicates, as well as between control and RNase experiments, confirmed a high degree of reproducibility at the global protein identification level (Supp. Fig. 2A–D). A manual examination of the overlap of the dataset revealed that many, but not all, high-confidence hits were shared between the two replicates (Fig. 1C, supp. Table 3). We therefore directly assessed the overlap between high-confidence hits from each individual replicate by comparing each with a defined set of known, annotated RNA associated proteins (Hentze et al., 2018). We found a considerable degree of recovery and reproducibility, specifically 58.5% of proteins in replicate 1 and 59% in replicate 2 were known RNA associated proteins (Fig. 1C, Supp. Fig. 2E).

To complement this analysis, we evaluated our dataset using additional annotation sources and observed consistent enrichment of RNA related annotations across replicates. For example, our high confidence DIF-FRAC hits from both replicates were highly enriched for GO annotations associated with RNA and RNA binding (Fig. 1D and E).

Additionally, an analysis of protein superfamilies using the InterPro database (Hunter et al., 2009) revealed that proteins with high confidence DIF-FRAC scores in both replicates were also highly enriched in proteins associated with RNA-binding, nucleic acid binding, or the ribosome (Supp. Fig. 2F and G).

Thus, despite relatively limited overlap between our two replicates, both replicates were equally effective at identifying RNA-associated proteins. We attribute the limited overlap between replicates to the heterogeneity of the differentiated vertebrate tissue examined here. By comparison, most high-throughput methods for identifying RNA-associated proteins have been applied only to single cell types, usually in cultured cells. A complete list of high confidence hits present in both replicates can be found in Supp. Table 4.

To complement these systematic analyses, we manually curated our high-scoring hits from *Xenopus* DIF-FRAC. We found, for example, that of the 60 highest scoring ribosomal proteins, 52 displayed DIF-FRAC *Z-scores* greater than our cutoff of 2.97, with 46 scoring *Z-score* > 3.5 (Fig. 2A–D). We hasten to note, however, that any specific threshold should be interpreted with care. Indeed, additional ribosomal proteins scored below this threshold in our dataset despite having a generally consistent elution profile shift similar to other ribosomal proteins (Fig. 2E). We observed DIF-FRAC scores to behave in a manner typical of computational prioritization methods, with a trade-off between enrichment and recall of RNA-associated proteins (Supp. Fig. 2H and I), allowing researchers to select score thresholds that weigh these factors as desired for a particular experiment.

Other high confidence hits also included known components of universal RNA processing machines, such as nucleolin (Ncl), the spliceosome (Sf3b3, Puf60), and stress granules (G3bp2) (Fig. 2F and G, Supp. Table 2). More importantly, we identified several RNA-binding proteins with known roles in early *Xenopus* embryos, such as Lin28a, Staufen1 (Stau1), and Cirbp (Faas et al., 2013; Peng et al., 2006; van Venrooy et al., 2008; Yoon and Mowry, 2004) (Fig. 2H and I). By contrast, known negative controls such as Vps35 and Cop9 signalosome subunits showed no elution shift after RNase treatments (Fig. 2J and K). Supp. Fig. 3 shows histograms of the background distributions used to calculate *Z-scores* for select proteins. Supp. Fig. 4 shows replicate elution profiles demonstrating reproducibility.

In our previous study using DIF-FRAC, we observed distinct classes of altered elution profiles for distinct RNA-associated proteins (Mallam et al., 2019), and we observed similar trends in our data from *Xenopus*. For example, Ncl displayed a clear mobility shift after RNase treatment, but the protein continued to elute in defined peaks (Fig. 2F), suggesting that it is stable in the absence of RNA. On the other hand, for ribosomal proteins and Lin28a, overall observed abundance was drastically reduced after RNase treatment (Fig. 2B–E, H), indicating reduced protein stability in the absence of RNA. Finally, the splicing factor, Puf60, displayed a clear *increase* in observed abundance (Fig. 2G), likely because it is insoluble when complexed with RNA. Thus, both systematic analysis and manual curation indicate that DIF-FRAC effectively identified RNA-associated proteins in *Xenopus*.

Importantly, our list of high scoring proteins also contained several proteins not previously known to associate with RNA, providing new insights. For example, our high confidence DIF-FRAC hits were enriched with the "Lactate dehydrogenase/glycoside hydrolase" superfamily (Supp. Fig. 2F), consistent with recent data suggesting frequent interaction of metabolic enzymes with RNA and suggesting 'moonlighting' functions (Castello et al., 2015). The "Coatomer/clathrin adaptor appendage" superfamily was also enriched in our high confidence DIF-FRAC hits (Supp. Fig. 2F), including for example AP-2 complex sub-units (Supp. Fig. 5B and C). This result is consistent with reports of RNA interactions with the related COPI complex (Todd et al., 2013).

Our dataset also contained a number of microtubule-associated proteins not previously known to be RNA-associated. Among these, Ccdc124/Lso2 was of interest because it is linked to both ribosomes and the mitotic spindle (Telkoparan et al., 2013; Wang et al., 2018). Several
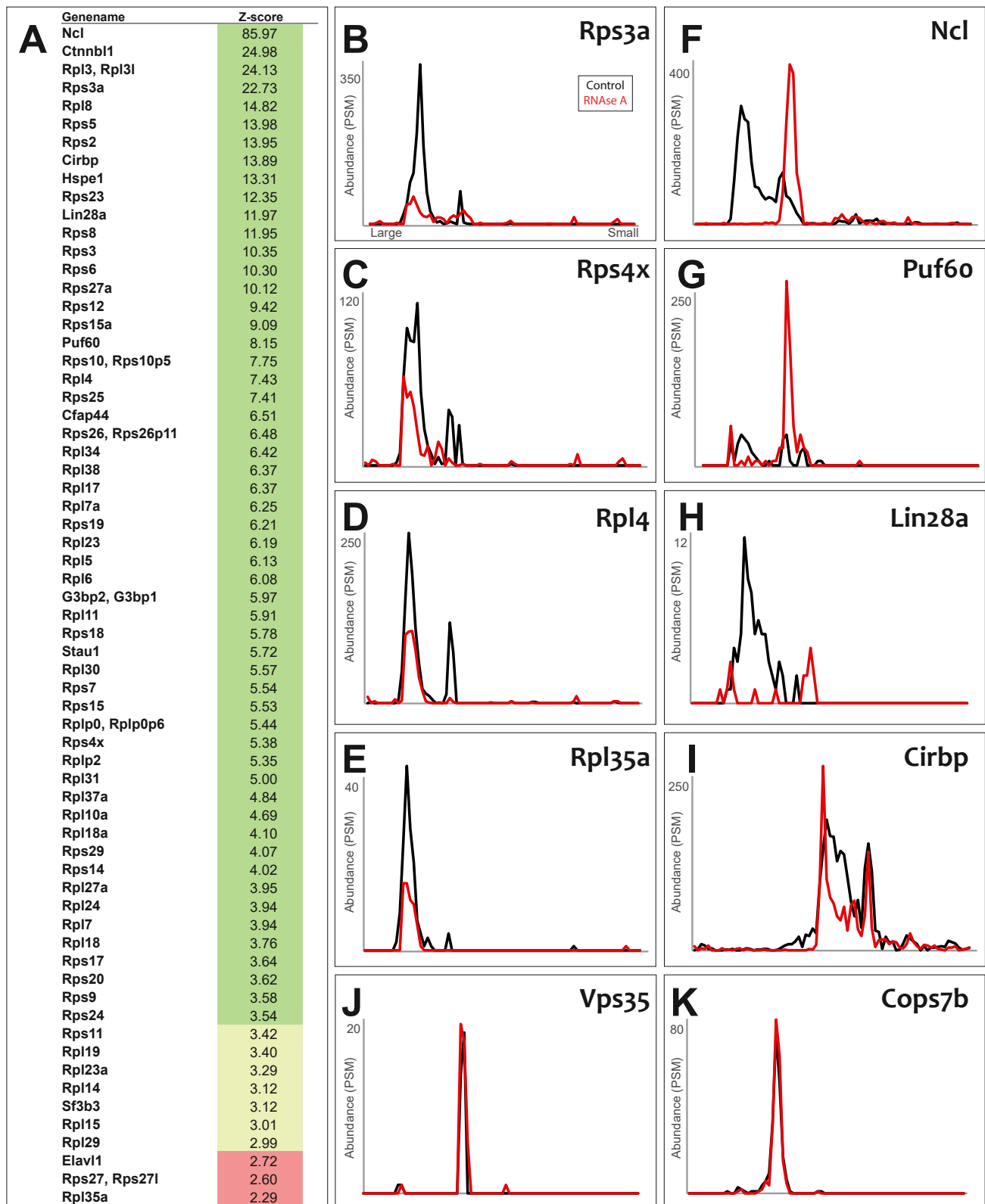
Fig. 2. Individual DIF-FRAC elution profiles show distinct changes consistent with RNAse sensitivity. (A) Table of DIF-FRAC calculated *Z-scores* for ribosomal proteins and selected others. Values highlighted in green and yellow are considered high confidence. Values in red are of borderline confidence and should be evaluated with prior knowledge. (B–D) Individual profiles for ribosomal subunits. X-axis represents SEC fractions from larger molecular weight to smaller. Y-axis represents observed abundance in MS by unique peptide spectral matches normalized to the highest value for that protein. (E) Ribosomal subunit, Rpl35a, had a *Z-score* below the cutoff (A), but its elution profile shows consistent behavior with other ribosomal subunits. (F) Known RNA-binding protein, Nucleolin, shows shift in molecular weight. (G) Known RNA-binding protein, Puf60, shows increased observed abundance. (H–I) Profiles of RNA-binding proteins with known roles in *Xenopus* development. (J–K) Elution profiles of negative controls do not change.

recent studies have explored spindle-associated RNAs (Blower et al., 2007; Jambhekar et al., 2014; Sharp et al., 2011), so it is notable that several additional spindle-associated proteins were identified here as RNA associated proteins. In both replicates of *Xenopus* DIF-FRAC, we identified Kif11/Eg2, Kif15, Map1s, Eml1, Eml2, and Map7d3 (Supp. Tables 2 and 3). These findings are of special interest in the context of multiciliated cells, since much of spindle machinery is shared with motile and primary cilia (Bernabé-Rubio et al., 2016; Smith et al., 2011).

We also identified several novel RNA-associated proteins that also have potential or known roles in ciliary beating. These include the axonemal dynein subunit Dnah11, the radial spoke protein Rsph1, and Hspe1, a heat shock family chaperone known to interact with the axonemal protein Spag16, which is also present in liquid-like nuclear speckles (Nagarkatti-Gude et al., 2011; Zhang et al., 2008) (Fig. 3A and B).

The most interesting novel RNA-associated protein identified here was Cfap44, encoded by a human ciliopathy gene and required for cilia motility in diverse organisms (Coutton et al., 2018; Fu et al., 2018; Kubo et al., 2018; Tang et al., 2017). The RNA-dependent change in elution profile for Cfap44 was unexpected, but we observed this elution shift for Cfap44 in both biological replicates (Supp. Fig. 4F). In both samples, Cfap44 displayed the distinct "gain in observed abundance" pattern described above for Puf60, which we interpret as a gain in protein solubility upon RNA degradation (Mallam et al., 2019).

In our previous work, incorporation of prior knowledge of protein-protein interactions allowed for identification of RNA-associated proteins that may not qualify as high confidence by their elution profiles alone (Mallam et al., 2019). This was also the case in our *Xenopus* DIF-FRAC (e.g. Rpl35a, Fig. 2A, E). This principle led us to then examine the elution profile of Cfap43, a known interaction partner of Cfap44 (Urbanska et al., 2018). Importantly, despite a relatively low DIF-FRAC *Z-score* (Fig. 3A), Cfap43 nonetheless displayed a clear shift upon RNase treatment that was strikingly similar to that of Cfap44 (Fig. 3D).

As an independent assessment of Cfap44 as a candidate RNA-associated protein, we used homology modeling to explore its structure. Using both pGenTHREADER ('CERT' confidence rating) and HHPred (Probability score: 99.83), we found that Cfap44 has strong similarity to the known RNA-associated protein Gemin5 (PDB ID: 5GXH) (Battle et al., 2006; Xu et al., 2016; Yong et al., 2010). We further found a WD40 domain between amino acids 400 and 1600 in Cfap44 (Fig. 3E) and using the existing co-crystal structure of Gemin5 bound to RNA (Xu et al., 2016), we were able to model an RNA molecule with our Cfap44 structural model (Fig. 3E). Using this structural model, we then interpreted amino acid sequence conservation data and observed a broad mixture of variable and conserved residues across the WD40 domain (Fig. 3e''). Focusing on residues near the predicted RNA binding pocket, however, we observed pockets of conserved residues (Fig. 3e'''). Although this model should be interpreted with care, these observations give additional credence to Cfap44's WD40 domain being RNA binding and provides testable residues to mutate for further investigations.

These results prompted us to consider what role an interaction with RNA may play in Cfap44 function. Because RNA is not thought to act in axonemes, we re-examined the localization of Cfap44 in more detail. Cfap44 localizes to the axonemes of *Chlamydomonas* flagella and *Tetrahymena* cilia (Coutton et al., 2018; Fu et al., 2018; Kubo et al., 2018; Urbanska et al., 2018), but its localization in vertebrate animals has not been reported. In *Xenopus* MCCs, we found that Cfap44 is strongly localized to ciliary axonemes (Fig. 4A).

Strikingly, we also observed strong GFP-Cfap44 localization to foci in the cytoplasm of MCCs (Fig. 4C), a pattern similar to that observed for DynAPs, MCC-specific liquid-like organelles that concentrate axonemal dyneins and their assembly factors (Huizar et al., 2018). Moreover, co-localization with the axonemal dynein assembly factor Ktu/Dnaaf2 revealed an even more interesting pattern; Cfap44+ foci are heterogeneous. Some Cfap44+ foci do NOT colocalize with DynAPs, while others partially co-localize (Fig. 4B-C). For Cfap44+ foci that do co-localize, the
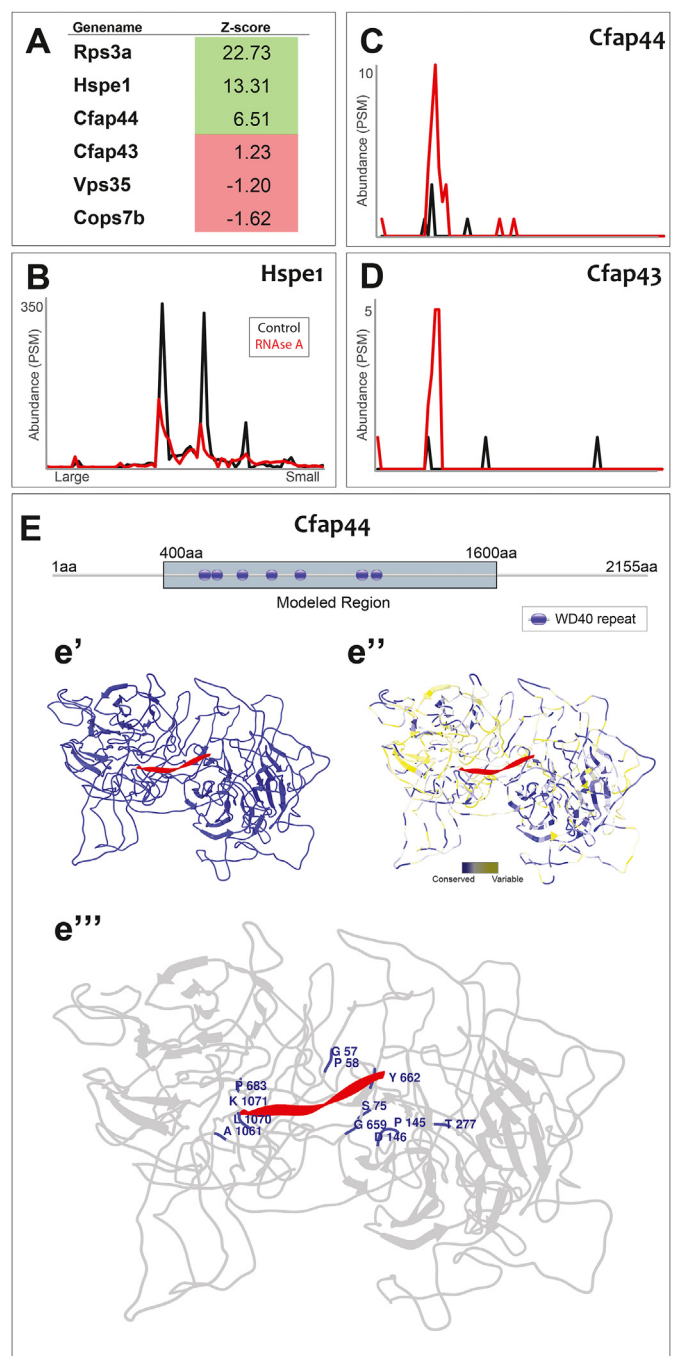

Fig. 3. DIF-FRAC identifies a ciliopathy protein as RNA associated. (A) Table of DIF-FRAC calculated *Z-scores* for selected motile cilia-related proteins; Rps3A, Vps35 and Cops7b serve as positive and negative controls. (B) Elution profile of Hspe1 shows loss of observed abundance. (C) Elution profile of the inner arm dynein tethering protein Cfap44 shows a gain of observed abundance. (D) Elution profile of Cfap43 shows similar behavior to Cfap44. (E) Graphic of modeled region of Cfap44 showing identified WD40 repeat segments. (e') Homology model of Cfap44 WD-40 domains (blue) with an RNA molecule (red) is modeled from Gemin5 crystal structure (PDB ID: 5GXH). (e'') Homology model of Cfap44 colored to show amino acid conservation where blue is highly conserved and yellow is variable. (e''') Homology model of Cfap44 highlighting highly conserved residues in proximity (<5.0 Å) to modeled RNA molecule.

correlation of signals is roughly similar to that observed recently for inner or outer arm dyneins that occupy functionally distinct sub-regions within DynAP (Lee et al., 2020). This result is consistent with the role of Cfap44 as a specific tether connecting only a specific subset of dynein arms (the dimeric, *f*-type inner arms) to axonemal microtubules (Fu et al., 2018; Kubo et al., 2018).

Finally, RNA-protein interactions are a common mechanism for assembly of liquid-like organelles (Banani et al., 2017; Shin and Brangwynne, 2017), so the localization of Cfap44 prompted us to ask if DynAPs contain RNA. Strikingly, staining of *Xenopus* MCCs with SYTO RNASelect revealed both the expected strong signal in nuclei, but also a substantial signal in DynAPs, as indicated by co-localization with Ktu (Fig. 4D; Supp. Fig. 6). Thus, DIF-FRAC identified Cfap44 as an RNA-associated protein, leading to our finding of Cfap44 localization to DynAPs, which in turn led us to discover that RNA is a novel component of DynAPs.

3. Discussion

The broad utility of large-scale rosters of RNA-associated proteins is now well-established, yet such resources for vertebrate embryonic tissues remain limited. Thus, both the data and the method reported here will be significant, because they will *a)* provide a valuable resource for understanding mucociliary epithelia and *b)* facilitate future RNA-associated protein discovery efforts in developing embryos.

The dataset we report provides new hypotheses that should lead to a better understanding of the interaction between microtubule-associated proteins and RNAs generally, as well as specific hypotheses concerning axonemal dynein assembly. Indeed, the only *axonemal* proteins previously found to be present in DynAPs are the dynein subunits themselves (Huizar et al., 2018). Our results therefore suggest the possibility that *f*-type inner arm dyneins may be pre-assembled together with their Cfap43/44 tether in the cytoplasm before deployment to the axoneme. This is in direct contrast to proteins that serve a similar function for the outer dynein arms, such as the "docking complex" proteins Ttc25, Armc4, and Mns1, which are *not* present in DynAPs (Huizar et al., 2018). Moreover, these results may shed light on the etiology of Cfap43/44-related human ciliopathies (e.g. (Coutton et al., 2018; Morimoto et al., 2019; Tang et al., 2017)).
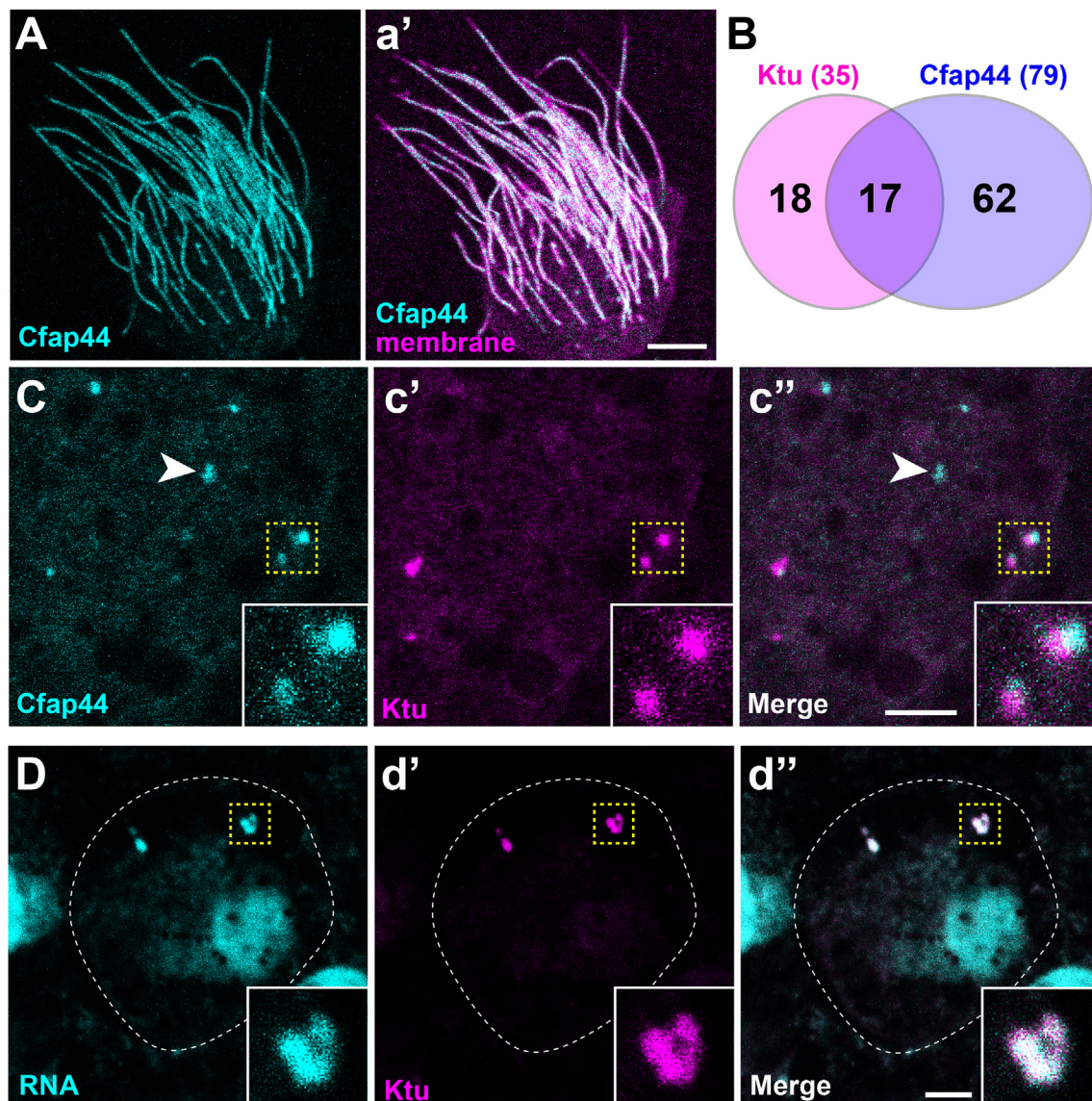

Fig. 4. Cfap44 and RNA are present in DynAPs: (A, a') GFP-Cfap44 localizes to axonemes in *Xenopus* motile cilia, as indicated by co-labelling with the membrane-RFP. (B) Overlap of Cfapp+ and Ktu + cytosolic foci in MCCs. (C, c', c'') GFP-Cfap44 labels cytosolic foci, some of which partially co-localize with DynAPs as indicated by co-labelling with mCherry-Ktu. This partial co-localization in DynAPs is reminiscent of that observed for inner or outer arm dynein subunits (Lee et al., 2020). (D, d', d'') Staining with SYTO RNASelect highlights RNA in the nucleus and also in DynAPs, as indicated by co-labelling with mCherry-Ktu.

More generally, these data should help us to better understand the connections between ubiquitous liquid-like organelles such as stress granules or P-bodies and cell type-specific organelles such as DynAPs. For example, stress granules and P-bodies share a large number of common components (Aizer et al., 2008; Jain et al., 2016). Likewise, DynAPs are enriched in broadly acting RNA-associated proteins, such as the heat shock chaperone Hsp90ab1 and the stress granule protein G3bp1 (Huizar et al., 2018), both of which were identified here as high confidence DIF-FRAC hits (Supp. Table 4). These findings add to the mounting evidence for a crucial role for RNA in DynAPs. For example, we also recently reported that the spliceosome subunit Sf3a3 physically interacts with inner dynein arm subunits and is also enriched in DynAPs (Lee et al., 2020); and axonemal dynein mRNAs have been reported to co-localize with chaperones in cytosolic foci in *Drosophila* sperm (Fingerhut and Yamashita, 2020). Thus, our data here provide an entry point for a deeper exploration of this interplay between broadly acting RNA-associated proteins and cell-type specific RNA-associated proteins such as Cfap44.

Perhaps the largest impact of this work is that it demonstrates the utility of DIF-FRAC for mapping RNA-associated proteomes in embryonic tissues. For example, the method could be readily adopted to explore the mechanisms of polarized mRNA localization in eggs, a ubiquitous developmental mechanism for which *Xenopus* has been a key model (Medioni et al., 2012; Sheets et al., 2017). Our success with animal caps differentiated into mucociliary epithelium also suggests the method will be effective for animal caps experimentally differentiated into any of a wide array of organs and tissues (e.g. (Ariizumi et al., 2009; Okabayashi and Asashima, 2003). Likewise, given the recent interest in studies of explanted tissues from zebrafish (e.g. (Williams and Solnica-Krezel, 2020; Xu et al., 2014), DIF-FRAC could also be readily applied to that powerful animal model.

An additional advance we report herein is the collapse of highly related proteins into orthogroups to substantially increase the total number of unique peptide spectral matches. We note here that our orthogroup-based method does not distinguish between homologous proteins that may have different capacity to associate with RNA. We anticipate this scenario to be rare but importantly, if one suspects members of an orthogroup to function differently with respect to RNA, the peptide level data is available from our dataset uploaded to PRIDE to distinguish differences in elution behavior among members of an orthogroup.

Ultimately, DIF-FRAC is an entirely label-free method and our use of orthogroups to collapse highly related proteins from multiple sources provides additional power, even when genomes are incompletely annotated or are non-diploid. As such, the findings here open the door to rapid, systematic identification of RNA-associated proteins in *any* model (or non-model) organism for which abundant tissue can be obtained.

4. Methods

Xenopus animal caps: Xenopus were housed and handled as described (Sive et al., 2000). Animal caps were dissected using forceps and cultured in 1X Steinberg's solution + gentamicin until sibling embryos reached stage 23 (Nieuwkoop and Faber, 1967). Explants were then lysed for 5 min on ice using 500 μl PierceIP Lysis Buffer (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) and dounced using pestle B. Lysate was then clarified 2x (13,000 g, 10 min, 4°C). Control lysate was left at room temperature for 30 min. RNAse lysate was treated with RNAse A (10μL, 100 μg, Thermo Fisher, catalog #EN0531) at room temperature for 30 min. Lysates were then filtered for multiple rounds (~5x) (0.45μm Ultrafree-MC filter unit (Millipore); 12,000 g, 2 min, 4°C) to remove insoluble aggregates. The remainder of the DIF-FRAC protocol is as previously described (Mallam et al., 2019) with modifications as discussed in the Results section above.

Chromatography: Both control and RNAse lysates were subjected to size exclusion chromatography (SEC) independently using a Thermo Scientific UltiMate 3000 HPLC system (Thermo Fisher Scientific, Waltham, MA). Soluble protein (200 μL at 10 mg/mL) was applied to a BioSep-SEC-s4000 600 × 7.8 mm ID, particle diameter 5 μm, pore diameter 500 Å (Phenomenex, Torrance, CA) and equilibrated in PBS, pH 7.2 at a flow rate of 0.5 mL min-1. Fractions were collected every 0.375 mL.

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 min, 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 desalted on a C18 filter plate (Part No. MFNSC18.10 Glygen Corp, Columbia, MD), 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 +2 or higher charge were selected for HCD 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 was acquired in the UT Austin Proteomics Facility.

Reference proteome construction: Two *Xenopus laevis* protein database were downloaded from Xenbase (http://www.xenbase.org/) (Karpinka et al., 2015; Nenni et al., 2019): the *X. laevis* JGI gene model (v9.1) peptide FASTA (http://ftp.xenbase.org/pub/Genomics/JGI/Xenla9.1/1.8.3.2/XL_9.1_v1.8.3.2.primaryTranscripts.pep.fa.gz) and an *X. laevis* protein sequence file containing GenBank sequences (http://ftp.xenbase.org/pub/Genomics/Sequences/xlaevisProtein.fasta). The files were filtered to remove sequences shorter than 20 amino acids or containing greater than 30% X's (*i.e.*, nonstandard amino acid symbol). A combined proteome was made by concatenating the JGI gene model FASTA (XL_9.1_v1.8.3.2.primaryTranscripts.pep.fa) and the GenBank sequence FASTA (xlaevisProtein.fasta). Each sequence in the combined file was mapped to eggNOG orthogroups using emapper (v 0.12.7)(Huerta-Cepas et al., 2017) and vertebrate-level Hidden Markov models downloaded from the eggNOG database in August 2017 (Huerta-Cepas et al., 2016). An orthology-collapsed proteome was then constructed by grouping protein sequences based on shared eggNOG orthology groups. Sequences were then concatenated separating each protein's sequence by triple lysines to ensure theoretical tryptic cleavage sites even when considering up to two missed tryptic cleavages. Proteins that did not map to any vertebrate-level eggNOG groups were retained in their original FASTA entry format.

Since *X. laevis* homologs and near-duplicate Xenbase/Genbank protein entries were combined into a single protein entry *via* the triple lysine concatenation, the orthology collapsed proteome is usable by any MS/MS search engine, which then processes the concatenated proteins properly into peptides (but for a single tailing lysine on the C-terminal-most peptide of internal proteins in the concatenation). Most significantly, by combining high sequence similarity proteins, the collapsed proteome substantially increases the rate of unique peptide assignment and overall protein identification (Supp. Table 1). After database spectral matching, entries described by eggNOG groups are related back to individual proteins *via* an annotation table mapping eggNOG IDs to eggNOG annotations, eggNOG IDs to Xenbase/Genbank IDs and XenBase/GenBank annotations, and eggNOG IDs to human UniProt IDs and human UniProt annotations.

The orthology collapsed proteome (XL9.1_XLgb_concat_orthocollapse_01102020.fasta) as well as accompanying scripts can be downloaded here: https://github.com/marcottelab/pivo.

Protein identification: Raw formatted mass spectrometry files were first converted to mzXML file format using MSConvert (http://proteo wizard.sourceforge.net/tools.shtml) and then processed using the MSBlender protein identification pipeline (Kwon et al., 2011), combining peptide-spectral matching scores from MSGF+ (Kim and Pevzner, 2014), X! TANDEM (Craig and Beavis, 2004) and Comet (Lingner et al., 2011) as peptide search engines with default settings. A false discovery rate of 1% was used for peptide identification. Elution profiles were assembled using unique peptide spectral matches for each eggNOG orthogroup across all fractions collected.

DIF-FRAC analysis: To identify proteins that are sensitive to RNAse treatment, we used our previously described statistical framework (Mallam et al., 2019). Briefly, each protein's control and RNAse elution profiles (Supp. Tables 5–8) are compared using a normalized Manhattan distance called the DIF-FRAC score to measure the overall change between profiles. Next, for each protein, we calculate a background distribution of DIF-FRAC scores. The background distribution is created from proteins with similar overall mass spec observed abundances to the target protein. More specifically, the distribution is made of a window of proteins with abundance rank +100 and −100 surrounding the target protein. Additionally, proteins with known RNA binding annotations are removed from the background distribution. These include proteins with GO "RNA Binding" annotations, Uniprot "ribonucleoprotein" annotations and proteins identified in high throughput RNA binding screens (Hentze et al., 2018). The distribution is then modeled using a two-component Gaussian mixture model where the component with the highest-mean represents RNA associated proteins, and the component with the lowest-mean represents non-RNA associated proteins (*i.e.,* background distribution). We then calculate a Z-score by comparing the DIF-FRAC score of the target protein to the background distribution. To specify a Z-score cutoff to be used for determining proteins as high confidence, we calculated for each Z-score a corresponding *p*-value and then false discovery corrected those p-values using Benjamini/Hochberg correction (Benjamini and Hochberg, 1995). We then applied a 0.05 *p*-value cutoff which resulted in Z-score cutoff values of >2.97 and > 3.3 for replicate 1 and replicate 2, respectively. We remind the reader that these cutoff values are ultimately arbitrarily chosen and are only a guide for determining our set of high confidence RNA associated proteins. We note that many additional RNA-associated proteins are likely to be found below these thresholds as well.

Systems-level analyses: For system-wide analyses in Fig. 1, eggNOG entries with high confidence DIF-FRAC scores were mapped to human Uniprot accessions. The gProfiler website (Raudvere et al., 2019) was used for Gene Ontology molecular function enrichment. Background proteins were defined to be the set of all identified proteins in the DIF-FRAC experiment. Electronic annotations were not considered in the analysis. Default values were used for all other parameters. For the InterPro superfamily enrichment analysis, Uniprot accessions were mapped to InterPro superfamily ids (Hunter et al., 2009) *via* Uniprot's ID mapping web service. Background distributions were calculated by taking random samples from all identified eggNOG entries. The occurrence of each InterPro superfamily was tabulated for each random sample and the mean and standard deviation were calculated across 1000 total random samplings.

Protein structure modelling: To model the structure of Cfap44, we used a template-based homology modeling approach. We first identified suitable templates for Cfap44 using default parameters from pGen-THREADER web server tool (Lobley et al., 2009) targeting amino acid sequence from position 400 to 1600 in Xenopus Cfap44 protein. We then used MODELLER (Sali and Blundell, 1993) to build the structure of Cfap44 using PDB ID: 2YMU as a template. 2YMU was the top hit of pGenTHREADER and had a 'CERT' confidence rating (Net Score: 273.346). pGenTHREADER also reported a Gemin5 structure (PDB ID: 5TEE) to have a 'CERT' confidence rating (Net Score: 191.361, Alignment Length: 604). We confirmed this hit using HHPred which scored Gemin5 with a probability of 99.83 over 566 alignment columns. 5TEE is an

apo-structure. We then chose an RNA bound structure of Gemin5 (PDB ID: 5GXH) for further comparison. The alignment of the Cfap44 structure model with Gemin5 was done using Chimera (Pettersen et al., 2004). ConSurf (Ashkenazy et al., 2016) was used with default settings to calculate residue conservation for Cfap44 and Chimera was used to visualize results.

Imaging: Full length of *Xenopus cfap44* was identified from Xenbase (Nenni et al., 2019), was amplified from *Xenopus* cDNA library and was cloned into pCS10R vector fused with N-terminal GFP. Capped cfap44 mRNA was synthesized using mMESSAGE mMACHINE SP6 transcription kit (ThermoFisher Scientific). Each 90 pg of GFP-cfap44 and mCherry-Dnaaf2/Ktu mRNAs were injected into two ventral blastomeres and live-imaging was performed as previously described (in Huizar et al., 2018). For RNA staining, embryos were fixed with Dent's fixative (20% DMSO and 80% methanol) at stage 23 and were stained by 500 nM of SYTO RNASelect Green Fluorescent Cell Stain (Invitrogen) for 20 min and were imaged after washing.

Data deposition: Proteomics data has been deposited in the PRIDE repository with accessions PXD017659 and PXD017650.

Code repository: Source code is freely available on GitHub: https://g ithub.com/marcottelab/diffrac.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ydbio.2020.08.008.

References

Aizer, A., Brody, Y., Ler, L.W., Sonenberg, N., Singer, R.H., Shav-Tal, Y., 2008. The dynamics of mammalian P body transport, assembly, and disassembly in vivo. Mol. Biol. Cell 19, 4154–4166.

Ariizumi, T., Takahashi, S., Chan, T.c., Ito, Y., Michiue, T., Asashima, M., 2009. Isolation and differentiation of Xenopus animal cap cells. Current protocols in stem cell biology 9, 5.1-1D. 5.31.

Artiles, K.L., Fire, A.Z., Frøkjær-Jensen, C., 2019. Assessment and maintenance of unigametic germline inheritance for C. elegans. Dev. Cell 48, 827–839 e829.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350.

Baltz, A.G., Munschauer, M., Schwanhausser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., Landthaler, M., 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell. 46, 674–690.

Banani, S.F., Lee, H.O., Hyman, A.A., Rosen, M.K., 2017. Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 18, 285–298.

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

Battle, D.J., Lau, C.K., Wan, L., Deng, H., Lotti, F., Dreyfuss, G., 2006. The Gemin5 protein of the SMN complex identifies snRNAs. Mol. Cell. 23, 273–279.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 289–300.

Bernabé-Rubio, M., Andrés, G., Casares-Arias, J., Fernández-Barrera, J., Rangel, L., Reglero-Real, N., Gershlick, D.C., Fernández, J.J., Millán, J., Correas, I., 2016. Novel role for the midbody in primary ciliogenesis by polarized epithelial cells. J. Cell Biol. 214, 259–273.

Blower, M.D., Feric, E., Weis, K., Heald, R., 2007. Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. J. Cell Biol. 179, 1365–1373.

Bolze, A., Mahlaoui, N., Byun, M., Turner, B., Trede, N., Ellis, S.R., Abhyankar, A., Itan, Y., Patin, E., Brebner, S., 2013. Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. Science 340, 976–978.

Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwinn, M.K., 2016. SONAR discovers RNA-binding proteins from analysis of large-scale protein-protein interactomes. Mol. Cell. 64, 282–293.

Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., Hentze, M.W., 2016. Comprehensive identification of RNA-binding domains in human cells. Mol. Cell. 63, 696–710.

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

Castello, A., Hentze, M.W., Preiss, T., 2015. Metabolic enzymes enjoying new partnerships as RNA-binding proteins. Trends Endocrinol. Metabol. 26, 746–757.

Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., Kettenbach, A.N., Diederichs, S., 2019. R-DeeP: proteome-wide and quantitative identification of RNA-dependent proteins by density gradient ultracentrifugation. Mol. Cell.

Cenik, E.S., Meng, X., Tang, N.H., Hall, R.N., Arribere, J.A., Cenik, C., Jin, Y., Fire, A., 2019. Maternal ribosomes are sufficient for tissue diversification during embryonic development in C. elegans. Dev. Cell 48, 811–826 e816.

Choesmel, V., Bacqueville, D., Rouquette, J., Noaillac-Depeyre, J., Fribourg, S., Crétien, A., Leblanc, T., Tchernia, G., Da Costa, L., Gleizes, P.-E., 2007. Impaired ribosome biogenesis in Diamond-Blackfan anemia. Blood 109, 1275–1283.

Chung, M.I., Kwon, T., Tu, F., Brooks, E.R., Gupta, R., Meyer, M., Baker, J.C., Marcotte, E.M., Wallingford, J.B., 2014. Coordinated genomic control of ciliogenesis and cell movement by RFX2. Elife 3, e01439.

Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.E., Ben Mustapha, S.F., Le Tanno, P., Wambergue-Legrand, C., Karaouzene, T., Martinez, G., Crouzy, S., Daneshipour, A., Hosseini, S.H., Mitchell, V., Halouani, L., Marrakchi, O., Makni, M., Latrous, H., Kharouf, M., Deleuze, J.F., Boland, A., Hennebicq, S., Satre, V., Jouk, P.S., Thierry-Mieg, N., Conne, B., Dacheux, D., Landrein, N., Schmitt, A., Stouvenel, L., Lores, P., El Khouri, E., Bottari, S.P., Faure, J., Wolf, J.P., Pernet-Gallay, K., Escoffier, J., Gourabi, H., Robinson, D.R., Nef, S., Dulioust, E., Zouari, R., Bonhivers, M., Toure, A., Arnoult, C., Ray, P.F., 2018. Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat. Commun. 9, 686.

Craig, R., Beavis, R.C., 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466–1467.

Faas, L., Warrander, F.C., Maguire, R., Ramsbottom, S.A., Quinn, D., Genever, P., Isaacs, H.V., 2013. Lin28 proteins are required for germ layer specification in Xenopus. Development 140, 976–986.

Fingerhut, J.M., Yamashita, Y.M., 2020. Localized mRNA translation mediates maturation of cytoplasmic cilia in Drosophila spermatogenesis. BioRxiv. https://doi.org/10.1101/2020.04.21.054247.

Fu, G., Wang, Q., Phan, N., Urbanska, P., Joachimiak, E., Lin, J., Wloga, D., Nicastro, D., 2018. The I1 dynein-associated tether and tether head complex is a conserved regulator of ciliary motility. Mol. Biol. Cell 29, 1048–1059.

Gerstberger, S., Hafner, M., Tuschl, T., 2014. A census of human RNA-binding proteins. Nat. Rev. Genet. 15, 829.

Haag, E.S., Dinman, J.D., 2019. Still searching for specialized ribosomes. Dev. Cell 48, 744–746.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., Bonasio, R., 2016. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. Mol. Cell. 64, 416–430.

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

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

Horani, A., Ustione, A., Huang, T., Firth, A.L., Pan, J., Gunsten, S.P., Haspel, J.A., Piston, D.W., Brody, S.L., 2018. Establishment of the early cilia preassembly protein complex during motile ciliogenesis. Proc. Natl. Acad. Sci. U.S.A. 115, E1221–E1228.

Huang, R., Han, M., Meng, L., Chen, X., 2018. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. Proc. Natl. Acad. Sci. Unit. States Am. 115, E3879–E3887.

Huerta-Cepas, Jaime, Forslund, Kristoffer, Pedro Coelho, Luis, Szklarczyk, Damian, Juhl Jensen, Lars, von Mering, Christian, Bork, Peer, 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. https://doi.org/10.1093/molbev/msx148.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286–D293.

Huizar, R.L., Lee, C., Boulgakov, A.A., Horani, A., Tu, F., Marcotte, E.M., Brody, S.L., Wallingford, J.B., 2018. A liquid-like organelle at the root of motile ciliopathy. Elife 7.

Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Mulder, N., Natale, D., Orengo, C., Quinn, A.F., Selengut, J.D., Sigrist, C.J., Thimma, M., Thomas, P.D., Valentin, F., Wilson, D., Wu, C.H., Yeats, C., 2009. InterPro: the integrative protein signature database. Nucleic Acids Res. 37, D211–D215.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., Parker, R., 2016. ATPase-modulated stress granules contain a diverse proteome and substructure. Cell 164, 487–498.

Jambhekar, A., Emerman, A.B., Schweidenback, C.T., Blower, M.D., 2014. RNA stimulates Aurora B kinase activity during mitosis. PloS One 9, e100748.

Karpinka, J.B., Fortriede, J.D., Burns, K.A., James-Zorn, C., Ponferrada, V.G., Lee, J., Karimi, K., Zorn, A.M., Pd, V., 2015. Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes. Nucleic Acids Res. 43, D756–D763.

Kim, S., Pevzner, P.A., 2014. MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. 5, 5277.

Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Xue, S., Ishijima, J., Shiroishi, T., Barna, M., 2011. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell 145, 383–397.

Kubo, T., Hou, Y., Cochran, D.A., Witman, G.B., Oda, T., 2018. A microtubule-dynein tethering complex regulates the axonemal inner dynein f (I1). Mol. Biol. Cell 29, 1060–1074.

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

Lee, C., Papoulas, O., Cox, R.M., Horani, A., Drew, K., Brody, S.L., Marcotte, E.M., Wallingford, J.B., 2020. DynAPs partition inner and outer arm dyneins into discrete sub-compartments containing specific regulatory factors. *BioRxiv*. https://doi.org/10.1101/2020.04.21.052837.

Li, Y., Zhao, L., Yuan, S., Zhang, J., Sun, Z., 2017. Axonemal dynein assembly requires the R2TP complex component Pontin. Development 144, 4684–4693.

Lin, Y., Protter, D.S., Rosen, M.K., 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., Meinicke, P., 2011. CoMet–a web server for comparative functional profiling of metagenomes. Nucleic Acids Res. 39, W518–W523.

Lobley, A., Sadowski, M.I., Jones, D.T., 2009. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics 25, 1761–1767.

Ma, L., Quigley, I., Omran, H., Kintner, C., 2014. Multicilin drives centriole biogenesis via E2f proteins. Genes Dev. 28, 1461–1471.

Mali, G.R., Yeyati, P.L., Mizuno, S., Dodd, D.O., Tennant, P.A., Keighren, M.A., Zur Lage, P., Shoemark, A., Garcia-Munoz, A., Shimada, A., Takeda, H., Edlich, F., Takahashi, S., von Kreigsheim, A., Jarman, A.P., Mill, P., 2018. ZMYND10 functions in a chaperone relay during axonemal dynein assembly. Elife 7.

Mallam, A.L., Sae-Lee, W., Schaub, J.M., Tu, F., Battenhouse, A., Jang, Y.J., Kim, J., Wallingford, J.B., Finkelstein, I.J., Marcotte, E.M., Drew, K., 2019. Systematic discovery of endogenous human ribonucleoprotein complexes. Cell Rep. 29, 1351–1368 e1355.

McWhite, C.D., Papoulas, O., Drew, K., Cox, R.M., June, V., Dong, O.X., Kwon, T., Wan, C., Salmi, M.L., Roux, S.J., Browning, K.S., Chen, Z.J., Ronald, P.C., Marcotte, E.M., 2020. A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell 181, 460–474 e414.

Medioni, C., Mowry, K.L., Besse, L., 2012. Principles and roles of mRNA localization in animal development. Development 139, 3263–3276.

Mittag, T., Parker, R., 2018. Multiple modes of protein-protein interactions Promote RNP granule assembly. J. Mol. Biol. 430, 4636–4649.

Morimoto, Y., Yoshida, S., Kinoshita, A., Satoh, C., Mishima, H., Yamaguchi, N., Matsuda, K., Sakaguchi, M., Tanaka, T., Komohara, Y., Imamura, A., Ozawa, H., Nakashima, M., Kurotaki, N., Kishino, T., Yoshiura, K.I., Ono, S., 2019. Nonsense mutation in CFAP43 causes normal-pressure hydrocephalus with ciliary abnormalities. Neurology 92, e2364–e2374.

Nagarkatti-Gude, D.R., Jaimez, R., Henderson, S.C., Teves, M.E., Zhang, Z., Strauss III, J.F., 2011. Spag16, an axonemal central apparatus gene, encodes a male germ cell nuclear speckle protein that regulates SPAG16 mRNA expression. PloS One 6, e20625.

Nenni, M.J., Fisher, M.E., James-Zorn, C., Pells, T.J., Ponferrada, V., Chu, S., Fortriede, J.D., Burns, K.A., Wang, Y., Lotay, V.S., 2019. Xenbase: facilitating the use of Xenopus to model human disease. Front. Physiol. 10, 154.

Nieuwkoop, P.D., Faber, J., 1967. Normal Table of Xenopus laevis. Garland, New York.

Okabayashi, K., Asashima, M., 2003. Tissue generation from amphibian animal caps. Curr. Opin. Genet. Dev. 13, 502–507.

Peng, Y., Yang, P.H., Tanner, J.A., Huang, J.D., Li, M., Lee, H.F., Xu, R.H., Kung, H.F., Lin, M.C., 2006. Cold-inducible RNA binding protein is required for the expression of adhesion molecules and embryonic cell movement in Xenopus laevis. Biochem. Biophys. Res. Commun. 344, 416–424.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Queiroz, R.M., Smith, T., Villanueva, E., Marti-Solano, M., Monti, M., Pizzinga, M., Mirea, D.-M., Ramakrishna, M., Harvey, R.F., Dezi, V., 2019. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). Nat. Biotechnol. 37, 169–178.

Quigley, I.K., Kintner, C., 2017. Rfx2 stabilizes Foxj1 binding at chromatin loops to enable multiciliated cell gene expression. PLoS Genet. 13, e1006538.

Quinn, J.J., Chang, H.Y., 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17, 47.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–w198.

Sahoo, P.K., Smith, D.S., Perrone-Bizzozero, N., Twiss, J.L., 2018. Axonal mRNA transport and translation at a glance. J. Cell Sci. 131, jcs196808.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

Sawyer, I.A., Sturgill, D., Sung, M.H., Hager, G.L., Dundr, M., 2016. Cajal body function in genome organization and transcriptome diversity. Bioessays 38, 1197–1208.

Session, A.M., Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., Takahashi, S., Fukui, A., Hikosaka, A., Suzuki, A., Kondo, M., van Heeringen, S.J., Quigley, I., Heinz, S., Ogino, H., Ochi, H., Hellsten, U., Lyons, J.B., Simakov, O., Putnam, N., Stites, J., Kuroki, Y., Tanaka, T., Michiue, T., Watanabe, M., Bogdanovic, O., Lister, R., Georgiou, G., Paranjpe, S.S., van Kruijsbergen, I., Shu, S., Carlson, J., Kinoshita, T., Ohta, Y., Mawaribuchi, S., Jenkins, J., Grimwood, J., Schmutz, J., Mitros, T., Mozaffari, S.V., Suzuki, Y., Haramoto, Y., Yamamoto, T.S., Takagi, C., Heald, R., Miller, K., Haudenschild, C., Kitzman, J., Nakayama, T., Izutsu, Y., Robert, J., Fortriede, J., Burns, K., Lotay, V., Karimi, K., Yasuoka, Y., Dichmann, D.S., Flajnik, M.F., Houston, D.W., Shendure, J., DuPasquier, L., Vize, P.D., Zorn, A.M., Ito, M., Marcotte, E.M., Wallingford, J.B., Ito, Y., Asashima, M., Ueno, N., Matsuda, Y., Veenstra, G.J., Fujiyama, A., Harland, R.M., Taira, M., Rokhsar, D.S., 2016. Genome evolution in the allotetraploid frog Xenopus laevis. Nature 538, 336–343.

Sharp, J.A., Plant, J.J., Ohsumi, T.K., Borowsky, M., Blower, M.D., 2011. Functional analysis of the microtubule-interacting transcriptome. Mol. Biol. Cell 22, 4312–4323.

Sheets, M.D., Fox, C.A., Dowdle, M.E., Blaser, S.I., Chung, A., Park, S., 2017. Controlling the Messenger: regulated translation of maternal mRNAs in Xenopus laevis development. In: Vertebrate Development. Springer, pp. 49–82.

Shin, Y., Brangwynne, C.P., 2017. Liquid phase condensation in cell physiology and disease. Science 357.

Sive, H.L., Grainger, R.M., Harland, R.M., 2000. Early Development of Xenopus laevis: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

Smith, K.R., Kieserman, E.K., Wang, P.I., Basten, S.G., Giles, R.H., Marcotte, E.M., Wallingford, J.B., 2011. A role for central spindle proteins in cilia structure and function. Cytoskeleton (Hoboken) 68, 112–124.

Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, J., Zhang, L., Sun, X., Zhi, E., Wang, H., Li, H., Jin, L., Luo, Y., Wang, J., Yang, S., Zhang, F., 2017. Biallelic mutations in CFAP43 and CFAP44 cause male infertility with multiple morphological abnormalities of the sperm flagella. Am. J. Hum. Genet. 100, 854–864.

Telkoparan, P., Erkek, S., Yaman, E., Alotaibi, H., Bayik, D., Tazebay, U.H., 2013. Coiled-coil domain containing protein 124 is a novel centrosome and midbody protein that interacts with the Ras-guanine nucleotide exchange factor 1B and is involved in cytokinesis. PloS One 8, e69289.

Todd, A.G., Lin, H., Ebert, A.D., Liu, Y., Androphy, E.J., 2013. COPI transport complexes bind to specific RNAs in neuronal cells. Hum. Mol. Genet. 22, 729–736.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., Meister, G., 2017. A compendium of RNA-binding proteins that regulate microRNA biogenesis. Mol. Cell. 66, 270–284 e213.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., Krijgsveld, J., 2019. The human RNA-binding proteome and its dynamics during translational arrest. Cell 176, 391–403 e319.

Urbanska, P., Joachimiak, E., Bazan, R., Fu, G., Poprzeczko, M., Fabczak, H., Nicastro, D., Wloga, D., 2018. Ciliary proteins Fap43 and Fap44 interact with each other and are essential for proper cilia and flagella beating. Cell. Mol. Life Sci. 75, 4479–4493.

van Venrooy, S., Fichtner, D., Kunz, M., Wedlich, D., Gradl, D., 2008. Cold-inducible RNA binding protein (CIRP), a novel XTcf-3 specific target gene regulates neural development in Xenopus. BMC Dev. Biol. 8, 77.

Walentek, P., Quigley, I.K., 2017. What we can learn from a tadpole about ciliopathies and airway diseases: using systems biology in Xenopus to study cilia and mucociliary epithelia. Genesis 55.

Wang, Y.J., Vaidyanathan, P.P., Rojas-Duran, M.F., Udeshi, N.D., Bartoli, K.M., Carr, S.A., Gilbert, W.V., 2018. Lso2 is a conserved ribosome-bound protein required for translational recovery in yeast. PLoS Biol. 16, e2005903.

Werner, M.E., Mitchell, B.J., 2012. Understanding ciliated epithelia: the power of Xenopus. Genesis 50, 176–185.

Wickramasinghe, V.O., Venkitaraman, A.R., 2016. RNA processing and genome stability: cause and consequence. Mol. Cell. 61, 496–505.

Williams, M.L., Solnica-Krezel, L., 2020. Nodal and planar cell polarity signaling cooperate to regulate zebrafish convergence and extension gastrulation movements. Elife 9.

Wühr, M., Freeman Jr., R.M., Presler, M., Horb, M.E., Peshkin, L., Gygi, S.P., Kirschner, M.W., 2014. Deep proteomics of the Xenopus laevis egg using an mRNA-derived reference database. Curr. Biol. 24, 1467–1475.

Xu, C., Ishikawa, H., Izumikawa, K., Li, L., He, H., Nobe, Y., Yamauchi, Y., Shahjee, H.M., Wu, X.H., Yu, Y.T., Isobe, T., Takahashi, N., Min, J., 2016. Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. Genes Dev. 30, 2376–2390.

Xu, P.-F., Houssin, N., Ferri-Lagneau, K.F., Thisse, B., Thisse, C., 2014. Construction of a vertebrate embryo from two opposing morphogen gradients. Science 344, 87–89.

Xue, S., Barna, M., 2012. Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. 13, 355.

Yong, J., Kasim, M., Bachorik, J.L., Wan, L., Dreyfuss, G., 2010. Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. Mol. Cell. 38, 551–562.

Yoon, Y.J., Mowry, K.L., 2004. Xenopus Staufen is a component of a ribonucleoprotein complex containing Vg1 RNA and kinesin. Development 131, 3035–3045.

Zhang, Z., Shen, X., Jones, B.H., Xu, B., Herr, J.C., Strauss III, J.F., 2008. Phosphorylation of mouse sperm axoneme central apparatus protein SPAG16L by a testis-specific kinase. TSSK2. *Biol. Reprod.* 79, 75–83.