Phase Separations Promoted by RNA Scaffolds and Protein Sequestration in Fragile X Tremor Ataxia Syndrome

Teresa Botta-Orfila[1,2], Fernando Cid-Samper[1,2], Mariona Gelabert-Baldrich[1,2], Benjamin Lang[1,2], Nieves Lorenzo-Gotor[1,2], Riccardo Delli Ponti[1,2], Lies-Anne WFM Severijnen[3], Benedetta Bolognesi[1,2], Ellen Gelpi[4,5], Renate K. Hukema[3] and Gian Gaetano Tartaglia[1,2,6]

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain

[2] Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain

[3] Department of Clinical Genetics, Erasmus MC, 3000 CA Rotterdam, The Netherlands

[4] Neurological Tissue Bank of the Biobanc-Hospital Clinic-IDIBAPS, Barcelona, Spain

[5] Institute of Neurology, Medical University of Vienna, Vienna, Austria

[6] Institució Catalana de Recerca i Estudis Avançats (ICREA), 23 Passeig Lluís Companys, 08010 Barcelona, Spain

* To whom correspondence should be addressed.
Tel: +34 93 316 01 16; Fax: +34 93 396 99 83; Email: gian.tartaglia@crg.eu

Keywords

Scaffolding RNA, CGG repeat expansion, FMR1 premutation, Fragile-X associated tremor ataxia syndrome (FXTAS), RNA aggregates, RNA binding proteins (RBP), TRA2A splicing regulator, Neurodegeneration

ABSTRACT

Ribonucleoprotein (RNP) granules are dense aggregations composed of RNA-binding proteins (RBPs) and a part of the transcriptome that is still uncharacterized. We performed a large-scale study of interaction networks and discovered that RBPs phase-separating in RNP granules share specific RNA partners with high structural content, long untranslated regions (UTRs) and nucleotide repeat expansions. Our analysis suggests that RNAs can promote formation of RNP granules by acting as scaffolds for protein assembly.

To experimentally validate our findings, we investigated the scaffolding ability of CGG repeats contained in the 5' UTR of *FMR1* transcript implicated in Fragile X-associated Tremor / Ataxia Syndrome (FXTAS). Using a novel high-throughput approach we identified RBPs recruited by *FMR1* 5' UTRs of different lengths. We employed human primary tissues as well as primate and mouse models of FXTAS to characterize protein sequestration in *FMR1* aggregates and focused on a previously unreported partner, TRA2A, which induces impairment of splicing in genes linked to mental retardation and intellectual disability.

SIGNIFICANCE

Proteins and RNAs phase-separate into granules under physiological (ribosome assembly and shuttling through neurons) and pathological (Amyotrophic Lateral Sclerosis and Fragile X Tremor/Ataxia Syndrome) conditions.

We report here that granule-forming proteins share more RNA targets than other proteins and form a dense network of interactions. RNAs interacting with granule-forming proteins have longer UTRs, are more structured and contain nucleotide repeats. Our study suggests that specific RNAs can act as the 'glue' of ribonucleoprotein assemblies.

To experimentally validate the results of our analysis, we investigated how nucleotide repeats contained in the 5' UTR of FMR1 are able to sequester proteins in intranuclear inclusions. We predicted protein interactions *in silico*, confirmed them *in vitro* and further characterized the ribonucleoprotein aggregates *in vivo*.

Introduction

Ribonucleoprotein (RNP) aggregates or granules are liquid-like cellular compartments composed of RNA-binding proteins (RBPs) and transcript (1, 2) that assemble through a process of phase separation and are in dynamic exchange with the surrounding environment (3). RNP granules such as processing bodies and stress granules (SGs) are evolutionarily conserved from yeast to human (4) and contain constitutive protein components such as G3BP1 (yeast: Nxt3), TIA1 (Pub1), and TIAR (Ngr1) (5). Several RBPs involved in physiological liquid-liquid phase separations are prone to form amyloid aggregates upon amino acid mutations (1, 6) that induce a transition from a liquid droplet to a solid phase (7, 8). This observation has led to the proposal that a liquid-to-solid phase transition is a mechanism of cellular toxicity (7) in diseases such as Amyotrophic Lateral Sclerosis (9) and Myotonic Dystrophy (10).

Recent evidence indicates that RBPs act as scaffolding elements promoting RNP granule assembly through protein-protein interactions (PPI) (11, 12), but protein-RNA interactions (PRI) may also play a role in granule formation. Indeed, a recent work based on RNP granules purification through G3BP1 pull-down indicate that 10% of the human transcripts can assemble into SGs (13). If distinct RNA species are present in SGs, a fraction of them could be involved in mediating RBP sequestration. We recently observed that different RNAs can act as scaffolds for RNP complexes (14), which led us to the hypothesis that some transcripts could be involved in granules assembly.

We here investigated the scaffolding potential of specific RNAs by computational and experimental means. Combining protein-RNA and protein-protein interaction networks revealed by enhanced CrossLinking and ImmunoPrecipitation (eCLIP) (15) and mass spectrometric analysis of SGs (4), we identified a class of transcripts that bind to a large number of proteins and, therefore, qualify as potential scaffolding elements. In agreement with recent literature reports (16), our calculations indicate that untranslated regions (UTRs) have a particularly strong potential to bind protein found in RNP granules, especially when they contain specific homo-nucleotide

repeats (17). In support of this observation, it has been reported for several diseases including Huntington's disease, Myotonic Dystrophy and several Ataxias that expanded trinucleotide regions trigger formation of intranuclear aggregates in which proteins are sequestered and inactivated (18, 19).

In a proof-of-concept series of experiments, we investigated the scaffolding ability of the 5'UTR of Fragile X Mental Retardation (*FMR1*) RNA. The 5′ UTR of *FMR1* RNA contains a CGG repeat that varies in length (the most common allele in Europe being of 30 repeats) and causes Fragile X syndrome at over 200 repeats, with methylation and silencing of the *FMR1* gene and lack of FMRP protein expression (20). The premutation range of *FMR1* (55–200 CGG repeats) is accompanied by appearance of *foci* (also called *FMR1* inclusions) that are the typical hallmark of Fragile X-associated Tremor / Ataxia Syndrome (FXTAS) (20). These *foci* are highly dynamic and behave as ribonucleoprotein granules (21) that phase separate in the nucleus forming inclusions (22).

Two mechanisms have been suggested to explain onset and development of FXTAS (23): i) RNA-mediated recruitment of proteins in *FMR1* inclusions and ii) aggregation of Repeat-Associated Non-AUG (RAN) polyglycines peptides translated from the *FMR1* 5' UTR (FMRpolyG) (20). Previous work indicates that *FMR1* inclusions contain specific proteins such as HNRNP A2/B1, MBNL1, LMNA and INA (24). Also FMRpolyG peptides (25) have been found in the inclusions, together with CUGBP1, KHDRBS1 and DGCR8 that are involved in splicing regulation, mRNA transport and regulation of microRNAs (26–29). Importantly, KHDRBS1 does not directly bind to *FMR1* (27), while its protein partner DGCR8 interacts physically with CGG repeats (28), which indicates that sequestration is a process led by a pool of protein drivers that progressively attract other networks.

The lability of *FMR1* inclusions, which makes them not suitable for biochemical purification (30, 31), does not allow to distinguish between physical and indirect partners of *FMR1* 5' UTR and there is need of novel strategies to characterize their interactome. Indeed, as the primary interactions are still largely unknown, current clinical strategies are limited to palliative care to ameliorate specific FXTAS

symptoms and there is still insufficient knowledge of targets for therapeutic intervention (20, 25) and a strong demand for new approaches (32).

Results and Discussions

We first investigated if RBPs aggregating in RNP granules interact with specific sets of proteins and RNAs. To discriminate proteins that are in RNP granules (granule RBPs) from other RBPs (non-granule RBPs) we relied on the most recent proteomics data on human and yeast SGs (**Materials and Methods** and **Supplementary Table 1A**). The protein-RNA interaction datasets were identified through eCLIP (human) (15) and microarray (yeast) (33) studies (**Supplementary Figure 1)**.

Protein-protein networks do not discriminate granule and non-granule RBPs

We analyzed if granule and non-granule RBPs show different interaction network properties. To this aim, we used available PPI datasets(33, 34) **(Materials and Methods)** (33, 34). We based our topological analysis on three centrality measures describing the importance of a node (protein) within the network. For each protein, we computed the degree (number of protein interactions), betweenness (number of paths between protein pairs) and closeness centrality (how close one protein is to other proteins in the network). We found that granule and non-granule RBPs networks display very similar topology both in yeast and human datasets **(Figure 1A; Supplementary Figure 2)**.

Protein-RNA networks robustly discriminate granule and non-granule RBPs

Having assessed that granule RBPs cannot be discriminated by properties of the PPIs network, we moved on to investigate PRIs of granule and non-granule RBPs. In both yeast and human, we found that PRIs increase significantly the centrality measures of the granule network **(Figure 1A** and **Supplementary Figure 2)**. Importantly, both yeast and human granule RBPs interact with more transcripts than other RBPs (**Figure 1B; Supplementary Figure 3; Supplementary Tables 1B, 1C, 1D** and **1E**; p-value yeast = 0.02, p-value human = 0.003, Wilcoxon test; **Materials and Methods**). Such a difference holds even when looking independently at either coding or non-coding RNAs (**Supplementary Figure 3; Supplementary Table 1D** and **1E;** p-value coding = 0.003, p-value non-coding = 0.01, Wilcoxon test) and upon

normalization by transcript length (p-value yeast = 0.02; p-value human = 0.002, Wilcoxon test).

Granule RBPs share RNA networks

We then wondered if granule RBPs interact with a common set of transcripts. For this analysis, we compared the number of shared transcripts between pairs of RBPs using the Jaccard index as a measure of target overlap. In both yeast and human we found that granule-forming proteins share a larger number of transcripts (**Figure 1C; Supplementary Tables 1F, 1G, 1H** and **1I;** p-value yeast < 2.2e-16, p-value K562 < 2.2e-16, KS test). Importantly, RNAs contacting at least one granule RBP preferentially interact with other granule RBPs (**Figure 1D**, p-value < 2.2e-16, Wilcoxon test). In agreement with this finding, RNAs interacting exclusively with granule RBPs have higher number of protein contacts than RNAs associating only with non-granule RBPs (**Supplementary Figure 4A,** p-value = 0.04, Wilcoxon test). This observation is consistent with a picture in which RNAs share a high number of RBPs interactions to promote recruitment into RNP granules. Using different confidence thresholds to select RBP partners (i.e., number of reads normalized by expression levels (35)), we found that our list of RNAs overlaps with a recently published atlas of transcripts enriched in SGs (Area under the ROC curve or AUC of 0.89; **Figure 1E**) (13).

Non-coding RNAs are contacted by granule RBPs

Among the most contacted RNAs we found an enrichment of small nuclear and nucleolar RNAs that are known to be associated with paraspeckles and Cajal bodies formation (**Supplementary Table 2A;** p-value < 2.2e-16, Wilcoxon test). We also identified a few highly contacted long non-coding RNAs such as NEAT1 that interacts with all the proteins present in our dataset (**Figure 1F**). In agreement with this finding, NEAT1 has been recently described as an architectural RNA implicated in scaffolding protein interactions (2) for paraspeckle nucleation (36). We hypothesize that other highly contacted long non-coding RNAs may have similar functions within cytoplasmic RNP granules. For instance, NORAD, a recently described long non-

coding RNA involved in genomic stability, interacts with all the proteins in our dataset (37). NORAD has repetitive sequence units, activation upon stress and ability to sequester proteins (38), which suggests a potential role in granule assembly.

Characteristic features of candidate scaffolding RNAs

We next analyzed which properties support the scaffolding activity of RNAs within granules. Comparing cases with the same total number of granule and non-granule interactions (**Materials and Methods**), we found that RNAs enriched in granule RBP contacts are more expressed (**Figure 2A**; p-value = 5e-11, KS test; **Materials and Methods**), structured (**Figures 2B**, and **2C** Parallel Analysis of RNA Structure PARS data; p-value = 0.005, KS test) and with longer UTRs (**Figure 2D**; p-value 5'UTR = 0.005, KS test; **Supplementary Figure 4B**). This result, also valid for yeast RNAs (**Supplementary Figures 4C** and **4D; Supplementary Tables 2B and 2C**), is consistent with previous observations that length (39), structure (40) and abundance (41) contribute to RNA assembly into RNP granules (13).

The increase in structural content is significant in the 5' UTRs of granule-associated transcripts (**Figure 2C**; p-value = 0.04; KS test) and nucleotide composition is specifically enriched in triplets with ability to assemble into hairpin structures (42): occurrence of CCG, UGC, CGC, GGU and CGG discriminate granule and non-granule transcripts (AUCs > 0.60; **Figure 2E**). In agreement with these findings, CROSS predictions of RNA structure (43) indicate that double-stranded regions are particularly enriched in granule-associated transcripts (**Supplementary Figure 5A**) and increase proportionally to the length of repeats (**Figure 2F**), which is in line with UV-monitored structure melting experiments (42).

In silico predictions indicate a large number of partners for scaffold RNAs

To further investigate the scaffolding ability of homo-nucleotide expansions, we selected the *FMR1* 5' UTR that contains CGG repetitions. Using *cat*RAPID *omics* (**Materials and Methods**) (44), we computed interactions between the 5' *FMR1* UTR (containing 79 CGG repeats) and a library of nucleic-acid binding proteins (3340 DNA-binding, RNA-binding and structurally disordered proteins (45)). Previously

identified CGG-binding proteins (27) such as HNRNP A1, A2/B1, A3, C, D and M, SRSF 1, 4, 5, 6, 7 and 10 as well as MBNL1 and KHDRBS3 were predicted to interact strongly (discriminative power > 0.90) and specifically (interaction strength > 0.90; **Figure 3A; Supplementary Table 3;** empirical p-values < 0.025). Similar binding propensities were also found for a set of 92 RBPs that have been previously reported to assemble in SGs (4) (**Supplementary Table 3**). In addition, our calculations identify a group of 37 RBP interactions that are predicted to form granules by the *cat*GRANULE algorithm (3) (**Materials and Methods; Supplementary Figures 5B and 5C**). Among the RBPs prone to associate with *FMR1*, we predict TRA2A (interaction score: 0.99; specificity 0.99; granule propensity = 2.15, i.e. the most granule-prone in our pool and ranking 188th out of 20190 human proteins).

High-throughput validation of CGG partners and identification of TRA2A interaction

We employed protein arrays to perform a large *in vitro* screening of RBP interactions with the first *FMR1* exon (46, 47). We probed both expanded (79 CGG) and normal (21 CGG) range repeats on independent replicas, obtaining highly reproducible results (Pearson's correlations >0.75 in log scale; **Figure 3B; Supplementary Table 4**). As a negative control, we used the 3' UTR of SNCA (575 nt) (47, 48).

Using fluorescence intensities (signal to background ratio) to measure binding affinities, we found that previously identified partners SRSF 1, 5 and 6 rank in the top 1% of all interactions (out of 8900 proteins), followed by KHDRBS3 (2%) and MBNL1 (5%). We observed strong intensities (signal to background ratio > 1.5 corresponding to top 1% of all interactions) for 85 RBPs interacting with expanded repeats (60 RBPs for normal-range repeats) and using stringent cut-offs (signal to background ratio > 2.5 or top 1 ‰ of all interactions) we identified 27 previously unreported interactions (binding to both expanded and normal range repeats).

The list of 85 RBPs showed enrichment in GO terms related to splicing activity (FDR $<10^{-7}$, as reported by GeneMANIA (49) and *clever*GO (50)) and includes SRSF 1, 5, 6 and 10, PCBP 1 and 2, HNRNP A0 and F, NOVA1, PPIG and TRA2A. *cat*RAPID

omics predictions are in strong agreement with protein array experiments: from low- to high-affinity interactions, *cat*RAPID performances increase progressively reaching AUCs of 0.80 (**Figure 3C**), which indicates strong predictive power. While KHDRBS1 (not present in the protein array) is predicted to not bind CGG repeats, two of its RBP partners, CIRBP and PTBP2, rank in the top 1% of all fluorescence intensities, as predicted by *cat*RAPID (48), and DGCR8, which interacts with KHDRBS1 through DROSHA (28), is also prone to interact (top 7% of all fluorescence intensities).

Out of 27 high-confidence candidates, 24 were predicted by *cat*GRANULE (3) to form granules and among them the splicing regulator TRA2A has the highest score (granule propensity = 2.15; **Figure 3D**; **Supplementary Figure 5D; Supplementary Table 3**). In agreement with our predictions, eCLIP experiments indicate that the *FMR1* transcript ranks in the top 25% of strongest interactions of TRA2A (15).

TRA2A recruitment in FMR1 inclusions is driven by CGG hairpins

As splicing defects have been reported to occur in FXTAS disease (23, 27), we decided to further investigate the sequestration of TRA2A. We first measured RNA and protein levels of TRA2A in B-lymphocytes of a normal individual (41 CGG repeats; Coriell repository number NA20244A) and a FXTAS premutation carrier (90 CGG repeats; Coriell repository number GM06906B). RNA and protein levels of TRA2A were found significantly increased 2.9 and 1.4 times in the FXTAS premutation carrier compared to normal individual, which indicates that the protein is significantly altered in disease (**Supplementary Figure 6**).

As nuclear inclusions do not form in B-lymphocytes, we used the COS-7 cell line to study *FMR1* inclusions (27). Transfection of a plasmid containing CGG expansions (triplet repeated 60 times) induce significant increase in RNA and protein levels of TRA2A after 48 hours (**Supplementary Figure 6**) (27). By means of RNA FISH coupled to immunofluorescence (**Materials and Methods**), we found that CGG expansions form nuclear inclusions that co-localize with endogenous TRA2A

(**Materials and Methods**) (27). TRA2A shows a diffuse nuclear pattern in cells that do not over-express CGG repeats (**Figure 4A**).

Upon knockdown of TRA2A using siRNA (**Materials and Methods**) we observed that the nuclear aggregates still form (**Figures 4B** and **4C**), while over-expression of TRA2A attached to GFP (GFP-TRA2A) result in strong recruitment within CGG inclusions (**Figure 4D**; control GFP plasmid and GFP-TRA2A in absence of CGG repeats does not give a granular pattern).

To further characterize the recruitment of TRA2A in CGG repeats, we treated COS-7 cells with two different chemicals. By incubating COS-7 cells with 9-hydroxy-5,11-dimethyl-2-(2-(piperidin-1-yl)ethyl)-6H-pyrido[4,3-b]carbazol-2-ium (also named *1a*) that binds to CGG repeats preventing interactions with RBPs (32), TRA2A sequestration was blocked (**Figure 5A**). Using TmPyP4 to specifically unfold CGG repeats (51) , we found that the aggregates are disrupted and TRA2A remains diffuse (**Figure 5B**). Our experiments show that the aggregation of TRA2A is caused by CGG repeats and depends on their structure.

TRA2A recruitment in RNA inclusions is independent of its partner TRA2B

Using RNA FISH coupled to immunofluorescence, we found that TRA2B, which is TRA2A partner, forms inclusions when COS-7 cells are transfected with CGG repeats, in agreement with previous *in vitro* screenings (**Figure 6A**) (26–29). The same result was observed upon TRA2A knockdown (**Figure 6B)**.

Importantly, endogenous TRA2A is still recruited by CGG inclusions upon TRA2B knockdown (**Figure 6B;** the result is also observed with over-expressed TRA2A; **Supplementary Figure 7**). By contrast, in absence of CGG, over-expressed TRA2A did not localize with inclusions (**Supplementary Figure 7A**). Yet, upon TRA2B knockdown and in presence of CGG repeats, over-expressed TRA2A is recruited in CGG inclusions (**Figure 6B and Supplementary Figure 7B**), which indicates that TRA2B does not mediate TRA2A sequestration.

Moreover, endogenous TRA2B co-localizes with TRA2A when both TRA2A and CGG are over-expressed (**Supplementary Figure 8**) and, similarly, endogenous TRA2A co-localizes with TRA2B when both TRA2B and CGG are over-expressed (**Supplementary Figure 8**). Upon TRA2A knockdown, there is co-localization of

over-expressed TRA2B with CGG aggregates. However, we observe a diffuse pattern in control cells upon over-expression of TRA2B with knockdown of TRA2A (**Supplementary Figure 9**).

Therefore, TRA2A and TRA2B are independently recruited by CGG RNA aggregates.

Effects of TRA2A recruitment on RNA splicing

To further investigate the functional implications of TRA2A recruitment in *FMR1* inclusions we analyzed changes in RNA splicing in COS-7 cells (**Materials and Methods**).

Splicing alterations due to TRA2A sequestration were studied through microarrays and RNA-seq experiments (both in triplicate experiments; **Materials and Methods**), which we used to investigate events (i) occurring upon CGG over-expression and (ii) affected when TRA2A is knocked-down. Applying these criteria, we identified 53 high-confidence exons subjected to splicing (36 skipped and 17 included exons; **Figure 7A; Supplementary Table 5; Material and Methods**). Using the *clever*GO algorithm (50) we found that the largest GO cluster of affected genes includes RBPs (18 genes; **Supplementary Table 5**; 'RNA-binding'; fold enrichment of 24; p-value $< 10^{-8}$; calculated with Bonferroni correction; examples: HNRNPL, CIRBP and DDX24) and, more specifically, spliceosome components ('mRNA splicing via spliceosome'; fold enrichment of 5; p-value $< 10^{-3}$, examples: ROA 2 and SRSF 10) or genes related to alternative splicing activity ('regulation of splicing'; fold enrichment of 6; p-value $< 10^{-3}$, examples: RBM5 and THOC1).

Intriguingly, genes associated with mental retardation, such as UBE2A (52), ACTB (53) and ACTG1 (54), have splicing patterns affected by TRA2A sequestration. Similarly, muscle related proteins, including PIP5K1C (55), TPM1 (56) and genes linked to intellectual disabilities such as DOCK3 (57) and craniofacial development, such as WWP (58, 59), are subjected to exon skipping upon TRA2A aggregation (**Supplementary Table 5; Figure 7B**). Out of 53 splicing events, 23 (including ACTG1, TMP1 and WWP2) involve transcripts that physically bind to FMRP protein,

as detected in other experiments (58), which indicates a new link (significance: p-value < 10^{-4}; Fisher's exact test) with Fragile X Syndrome (61).

Our results indicate that TRA2A recruitment in *FMR1* inclusions affects the splicing of a number of RBPs and genes associated with neuronal and muscular pathways.

TRA2A is present in murine and human FXTAS inclusions

Repeat associated non-AUG (RAN) translation has been shown to occur in *FMR1* 5'UTR, resulting in the production of FMRpolyG and FMRpolyA peptides (20). The main RAN translation product, FMRpolyG, co-localizes with ubiquitin in intranuclear inclusions, as observed in FXTAS human brain and mouse models (20, 25, 62).

We tested if TRA2A co-aggregates with FXTAS inclusions and FMRpolyG in a mouse model in which the *FMR1* 5'UTR (containing 90 CGG repeats) was expressed under the control of doxycycline (63). Immunohistochemistry experiments with sections of paraffin-embedded neurons and astrocytes show that TRA2A protein is indeed present in the inclusions (**Figure 8A**).

Remarkably, nuclear inclusions from FXTAS *post mortem* human brain donors show positive staining for TRA2A (**Figure 8B**) that also co-localize with FMRpolyG (**Figure 8C**).

Thus, TRA2A sequestration by CGG repeats is not only observed in cell lines, but also in FXTAS animal models and human *post mortem* brain samples.

Conclusions

In this work we investigated the role played by RNA in the assembly of RNP granules. Previous evidence indicates that proteins are the main cohesive elements within RNP granules: Banani and coworkers, for instance, showed that specific PPIs drive the formation of membrane-less compartments (64). Yet, based on our analysis of interaction networks, we propose that specific RNAs can serve as structural scaffolds to assemble proteins in RNP granules. According to our calculations, scaffolding RNAs are characterized by a large number of RBP contacts, increased length and structural content.

To experimentally validate our findings we focused on CGG repeat expansions present in the *FMR1* 5' UTR that trigger formation of RNP *foci* associated with the neurodegenerative disorder FXTAS (23, 27). Combining *in silico* predictions of protein-RNA interactions with large-scale *in vitro* validations (46, 47), we unveiled the interactome of *FMR1* 5' UTR, recovering previously reported interactions relevant in FXTAS such as SRSF1, 5, 6, KHDRBS3 and MBNL1, and identifying additional partners involved in alternative splicing, among which PCBP 1 and 2, HNRNP A0 and F, NOVA1, PPIG and TRA2A. In particular, we found that CGG repeats have a strong propensity to bind MBNL1, HNRNP A1, A2/B1, A3, C, D and M that have been previously shown to co-localize with *FMR1* inclusions. In agreement with other experimental reports (27), KHDRBS1 shows poor binding propensity to CGG repeats, while its RBP partners, CIRBP, PTBP2 and DGCR8 have stronger interactions, thus explaining its co-localization *in vivo* (28). At the time of writing, TRA2A has been reported to be one specific component of granules associated with Amyotrophic Lateral Sclerosis (65), which highlights an important link with neuropathology.

Using human primary tissues as well as primate and mouse models of FXTAS, we characterized TRA2A sequestration in *FMR1* inclusions. TRA2A is a splicing regulator containing Ser/Arg repeats near its N- and C-*termini* (66). Similarly to other splicing factors, TRA2A contains structurally disordered regions and multiple stretches of low sequence complexity (67) that promote liquid-liquid phase separation

(3). Although formation of large assemblies is crucial for splicing activity and regulation (68, 69), liquid-liquid phase separations have been also observed in the context of disease. For instance, aggregation of DMPF and ZNF9 transcripts (containing CUG or CCUG expansions) has been reported to induce sequestration of MBNL1 and partners (10) leading to impairment of splicing activity and Myotonic Dystrophy (70).

In a cellular model of FXTAS, we found that TRA2A is sequestered by nuclear RNA inclusions upon over-expression of *FMR1* 5' UTR. PPI networks do not play a direct role in protein recruitment, as identified in our *in silico* analysis: both endogenous and overexpressed TRA2A are recruited to CGG granules and upon knockdown of its partner TRA2B (71) the assembly process is still driven by the RNA scaffold. Unfolding the structure of CGG repeats through TMP4yP (51) or using 9-hydroxy-5,11-dimethyl-2-(2-(piperidin-1-yl)ethyl)-6H-pyrido[4,3-b]carbazol-2-ium to block protein binding (32) results in preventing TRA2A sequestration. Through splicing microarrays and RNA-seq analysis we found that TRA2A sequestration induces changes in alternative splicing of genes associated with mental retardation, including ACTB (51) and ACTG1 (52), intellectual disabilities, such as DOCK3 (55) and craniofacial development, such as WWP2 (56, 57), which are relevant in the context of Fragile X Syndrome (72).

We further tested the relevance of our findings in a FXTAS mouse model where we observed that TRA2A co-localizes with brain CGG inclusions both in neurons and astrocytes, thus providing evidence of involvement in disease. Supporting our findings, positive staining of TRA2A in inclusions from human *post mortem* brain donors with FXTAS indicates that TRA2A sequestration is linked with pathology. Although the non-AUG codon downstream the 5' UTR of *FMR1* (20) was not present in our *FMR1* constructs, when using brain samples from FXTAS post-mortem donors we observed that TRA2A co-localizes with the polypeptide FMRpolyG in brain inclusions. Thus, our data suggest that the 5' UTR of *FMR1* could also act as a scaffold for FMRpolyG. This result is in line with previous work in the field showing that interactions between proteins and cognate RNAs are frequent for aggregation-prone genes (73, 74).

Our work highlights general features that distinguish scaffolding RNAs present in the ribonucleoprotein granules and are not only specific to *FMR1* inclusions: the number of protein contacts and occurrence of nucleotide repeats as well as increase in secondary structure content and UTR length. In agreement with these findings, it has also been recently reported that nucleotide repeats are key elements promoting phase transitions (41). Moreover, two articles published at the time of writing reported that structural content (2) and UTR length (13) and are important properties of RNAs present in the granules. Importantly, our analysis shows that the abundance of molecules promoting phase separation is high, which has been previously shown to be important for protein aggregation (75–77) but also applies to RNA scaffolds.

Materials and Methods

Network analysis

We used the Jaccard index (**Supplementary Tables 1F, 1G, 1H** and **1I**) as a measure of the overlap of RNA targets between pairs of proteins. The Jaccard index of a specific couple of proteins a and b ($J_{a,b}$) was computed as:

$$J_{a,b} = \frac{|A \cap B|}{|A \cup B|}$$

were A is the set of RNA targets of the first protein of the pair, B is the set of RNA targets of the second protein of the pair.

RNA properties analysis

To study features of granule and non-granule transcripts, we compared RNAs with the same number of total proteins contacts (**Supplementary Table 1A**). We used transcripts with a total of 10 RBP contacts in the human analysis and 5 contacts in the yeast case. The number of RBP contacts was chosen to have granule and non-granule sets of comparable size (**Supplementary Tables 2B** and **2C**).

Additional details are provided in **Supporting Information.**

catRAPID omics analysis

*cat*RAPID *omics* was used to compute the interaction propensity of CGG repeats with proteins (44). *cat*RAPID *omics* ranks predictions based on interactions score as well as presence of motifs and RBDs (78). For RNA sequences > 1000 nt, the uniform fragmentation procedure is applied to determine the binding regions of a protein.

catGRANULE analysis

Structural disorder, nucleic acid binding propensity and amino acid patterns such as arginine-glycine and phenylalanine-glycine are key features of proteins coalescing in granules (3). These features were combined in a computational approach, *cat*GRANULE, that we employed to identify RBPs assembling into granules (scores > 0 indicate granule propensity) (3).

RNA IVT and Protein arrays

FMR1 5'-UTR expanded and control pCAGIG vectors, with 79 and 21 repeats, respectively, were generated by Dr. Marti's group (79). The UTRs were subcloned in PBSK plasmid containing promoters suitable for in vitro transcription. 50 pmoles of labeled RNA were hybridized in the protein arrays Human Protein Microarrays v5,2, Life Technologies. The arrays were dried and immediately scanned at 635nm in Microarray Scanner G2505B (Agilent). GenePix Pro 6.1 software (Molecular Devices) was used to determine the signal at 635nm of each spotted protein location and therefore quantify the RNA-protein interaction. Specifically, the local background intensity (B635) was subtracted from the intensity (F635) at each of the duplicate spots for a given protein, to quantify. Data was filtered based on signal to background ratio for each of the duplicate feature to be greater than 2.5 fold and Z-Score ≥ 3 from the global mean signal from all the spotted proteins. Finally, the intersection of technical replicates was considered as the final value for quantification.

Additional details are provided in **Supporting Information.**

IF-RNA FISH in COS-7 cells

COS-7 cells were grown on 13mm coverslips until a 70% confluence. Cells were transfected with lipofectamine 2000 (Invitrogen, #13778150) according to manufacturer's instructions and stained after 24 hours. siRNA treatments were performed with lipofectamine 2000 (Invitrogen, 11668019) transfection of (siTRA2A, Ambion AM16704, siTRA2B, Ambion S12749, compared to water). Overexpression of proteins was achieved by transfection of GFP-TRA2A, GFP-TRA2B, compared with GFP-vector, all plasmids kindly given by Dr. Nicolas

Charlet-Berguerand. For further treatments, cells were incubated with 25nM TmPyP4 (Abcam ab120793) or 15uM of 1a molecule (this latter kindly given by Matthew D. Disney).

Additional details are provided in **Supporting Information.**

Splicing Arrays experiments and analysis

COS-7 cells were grown in P10 plates and cultured in different conditions and in three biological replicates each: control, (CGG)60X 185ng, siTRA2A 50nM, (CGG)60X 185ng+siTRA2A 13.6ng, GFP-TRA2A 200ng, (CGG)60X 185ng + GFP-TRA2A 200ng. Total RNA extraction was performed with Qiagen RNeasy Mini Kit including DNAse treatment according to manufacturer's instructions. RNA amount was quantified and controlled with Nanodrop and Bioanalyzer. 100ng of total RNA from each sample were labelled according to the Affymetrix GeneChip® Whole Transcript Plus protocol, and hybridized to Affymetrix Human Clariom D array using a Affymetrix GeneChip Hybridization Oven 645, in Servei de Microarrays (IMIM-Barcelona).

Additional details are provided in **Supporting Information.**

RNA-seq experiments and analysis

An aliquot of the RNA extracted from COS-7 cells for splicing arrays (previous section) was used for RNASEQ, in three biological replicates each: control, (CGG)60X 185ng, siTRA2A 50nM, (CGG)60X 185ng+siTRA2A 13.6ng, GFP-TRA2A 200ng, (CGG)60X 185ng + GFP-TRA2A 200ng. 500ng of total RNA were used for library preparation with TruSeq total RNA-rRNA depletion (Illumina). The sequencing was performed with HiSeq3000, paired-end, 2X125, 3samples/lane.

Additional details are provided in **Supporting Information.**

ACKNOWLEDGEMENTS

The authors are deeply in debt to skin and brain donors and their relatives. We acknowledge specimen donations from Dr Nicolas Charlet-Berguerand (pcDNA3 CGG 60X), Dr David Elliott and Dr Caroline Dalgliesh (plasmid GFP TRA2A), Dr Eulalia Marti (plasmids CGG 21X and 79X and isolated fibroblasts from CAG carriers), Dr Matthew D. Disney (molecule *1a*). We thank Dr Fatima Gebauer, Dr Davide Cirillo and Dr Domenica Marchese for stimulating discussions.

The research leading to these results has been supported by European Research Council (RIBOMYLOME_309545), Spanish Ministry of Economy and Competitiveness (BFU2014-55054-P and BFU2017-86970-P) and "Fundació La Marató de TV3" (PI043296). We acknowledge support of the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013-2017'. We acknowledge the support of the CERCA Programme / Generalitat de Catalunya. Support of Spanish Ministry for Science and Competitiveness (MINECO) to the EMBL partnership.

Author contributions

GGT, BB and TBO conceived the study, FC performed the calculations, TBO, IGB, RH, LWS and EG performed the experiments. TBO, NLG, GGT and BL analyzed the data. TBO, FC, BB and GGT wrote the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Figures

Figure 1. RNA as a key element in the granule RBP network. A) The heat-map shows statistical differences between granule and non-granule elements (proteins or RNA) in protein-protein and protein-RNA networks. Only when including RNA interactions, granule and non-granule networks show different topologies (**Supplementary Figure 2**). **B)** Granule-forming RBPs show a high number of targets (K562 cell line, p-value = 0.003, Wilcoxon test); **C)** Granule-RBPs share more targets that other RBPs (p-value K562 < 2.2e-16, p-value yeast < 2.2e-16, KS test). **D)** RNAs contacting at least one granule RBP make more contacts with other RBP in the same granule network (p-value < 2.2e-16, Wilcoxon test). **E)** RNAs contacted by granule forming RBPs are enriched in SGs (4) (Area under the ROC curve AUC is used to measure the enrichment in SGs; **Materials and Methods**). Thresholds are the first (Q1), second (Q2) and third (Q3) quartile of the statistical distribution computed using number of reads normalized by expression levels. **F)** Protein contacts of non-coding RNAs at different confidence levels. Even for stringent thresholds, highly contacted transcripts are enriched in small nuclear (snRNAs) and small nucleolar RNAs (snoRNAs) (p-value < 2.2e-16, Wilcoxon test). Already described scaffolding RNAs such as NEAT1 are also highly contacted**.**

Figure 2. Properties of scaffolding RNAs. A), B), C) and **D)** Properties of RNAs contacted by granule-proteins. Granule transcripts are more abundant (**A**, p-value = 4.65e-11, KS test), structured (**B** and **C,** p-value = 0.005 and p-value=0.04; KS test) with longer UTRs (**D**, p-value 5'UTR = 0.005, KS test) than non-granule RNAs. **E)** Occurrence of CCG, UGC, CGC, GGU and CGG repeats discriminate the 5' UTRs of granule and non-granule transcripts (the Area under the ROC Curve AUC is used to separate the two groups). **F)** Increasing the length of CGG repeats results in stronger secondary structural content (the CROSS algorithm (43) is employed to measure the amount of double-stranded RNA).

Figure 3. Protein interactions of CGG repeats. A) Using *cat*RAPID *omics* (44), we computed protein interactions with the first *FMR1* exon (containing 79 CGG repeats). Previously identified partners, such as ROA 1, 2 and 3, HNRNP C, D and M, SRSF 1,

4, 5, 6, 7 and 10 as well as MML1 and KHDRBS3 show strong binding propensities and specificities (blue dots) (27). A previously unknown interactor, TRA2A (red dot) shows comparable binding propensities. **B)** We validated RBP interactions with FMR1 exon ("pre"containing 79 CGG repeats) through protein arrays (46). We obtained high reproducibility between replicas (Pearson's correlations > 0.75 in log scale) and identified strong-affinity interactions (signal to background ratio > 2.5; red dots). The same procedure was applied to FMR1 exon containing 21 CGG repeats (**Supplementary Table 4**). **C)** We measured *cat*RAPID *omics* (44) performances on protein array data (79 and 21 CGG repeats) selecting an equal number of strong- (highest signal to background ratios) and poor-affinity (lowest signal to background ratios) candidates. The performances increase by applying stringent cut-offs, reaching the AUC of 0.80. **D)** Out of 27 candidates binding to both 79 and 21 CGG repeats (signal to background ratio > 2.5), 15 are highly prone to form granules (blue bars) (3) and the splicing regulator TRA2A (red bar) shows the highest propensity. The black bars indicate non-specific partners (interacting also with SNCA 3' UTR (46) or showing poor RNA-binding propensity (45)).

Figure 4. Endogenous TRA2A is recruited in nuclear RNA inclusions upon CGG over-expression. This specific recruitment is validated by experiments with TRA2A over-expression and TRA2A knockdown. **A)** COS-7 cells were transfected with either CGG(60X) or the empty vector as control. After 24h of transfection cells were immunostained with primary antiTRA2A antibody and secondary 488 and hybridized with a Cy3-GGC(8X) probe for RNA FISH. The graph represents the 488/Cy3 intensities co-localization in the section from the white line. **B)** After 24h of transfection cells were immunostained with antiTRA2A antibody and hybridized with a Cy3-GGC(8X) probe for RNA FISH. **C)** Relative TRA2A protein levels in COS-7 cells treated as in B. D. COS-7 cells were transfected with empty vector or CGG(60X) and GFP-TRA2A. After 48h, cells were hybridized with Cy3-GGC(8X) probe for RNA FISH. The graph represents the GFP/Cy3 intensities co-localization in the section from the white line.

Figure 5. Disrupting CGG hairpins and dissolving RNA inclusions impair TRA2A sequestration. B) COS-7 cells were co-transfected with empty vector or CGG(60X), and after 24h of transfection cells were treated with 1a to block protein

binding. **B)** COS-7 cells were treated similarly as in **A)** but with TmPyP4 molecule instead of 1a to disrupt CGG structure. In both cases, cells were immunostained with primary anti TRA2A antibody and hybridized with Cy3-GGC(8X) probe for RNA FISH.

Figure 6. Endogenous TRA2B is recruited in CGG inclusions but TRA2A recruitment is independent from TRA2B. A) COS-7 cells were transfected with CGG(60X). After 24h of transfection cells were immunostained with antiTRA2B antibody and hybridized with a Cy3-GGC(8X) probe for RNA FISH. **B)** COS-7 cells were transfected with CGG(60X) and siTRA2A or siTRA2B. After 24h of transfection cells were immunostained with antiTRA2 antibodies and hybridized with a Cy3-GGC probe for RNA FISH. C. Relative TRA2A protein levels in COS-7 cells treated as in **A**.

Figure 7. TRA2A recruitment in *FMR1* inclusions affects alternative splicing. A) Splicing alterations due to TRA2A sequestration in COS-7 cells were studied using microarrays and RNA-seq. We identified 53 high-confidence genes subjected to exon skipping or inclusion (black: splicing events caused by CGG over-expression; red: splice events caused by CGG over-expression and altered upon TRA2A knocked-down; control: no CGG over-expression; fold changes and standard deviations are relative to the control; **Supplementary Table 5; Materials and Methods**). **Inset**: GO analysis of the largest cluster (18 genes) affected by TRA2A sequestration(50) reveals links to RBP activities and splicing activities. **B)** Exon inclusion/skipping of genes associated with mental retardation, such ACTB (51) and ACTG1 (52), intellectual disabilities, including DOCK3 (55) and craniofacial development WWP2 (56, 57) are shown (**Supplementary Table 5**).

Figure 8. TRA2A accumulations are present in murine and human FXTAS inclusions. A) TRA2A immunohistochemistry in wild type (WT) and premutated mouse model (counterstaining is done with haematoxylin). **B)** TRA2A immunohistochemistry in human hippocampus from control and FXTAS (counterstained with haematoxylin; the arrows points to the inclusion). **C)** Double immunofluorescence of TRA2A as well as FMRpolyG peptides in human FXTAS (**Materials and Methods**).

1. Hyman AA, Weber CA, Jülicher F (2014) Liquid-Liquid Phase Separation in Biology. *Annu Rev Cell Dev Biol* 30(1):39–58.
2. Maharana S, et al. (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*. doi:10.1126/science.aar7366.
3. Bolognesi B, et al. (2016) A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep* 16(1):222–231.
4. Jain S, et al. (2016) ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* 164(3):487–498.
5. Buchan JR, Muhlrad D, Parker R (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. *J Cell Biol* 183(3):441–455.
6. Kato M, et al. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149(4):753–767.
7. Patel A, et al. (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162(5):1066–1077.
8. Qamar S, et al. (2018) FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* 173(3):720-734.e15.
9. Murakami T, et al. (2015) ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* 88(4):678–690.
10. Pettersson OJ, Aagaard L, Jensen TG, Damgaard CK (2015) Molecular mechanisms in DM1 — a focus on foci. *Nucleic Acids Res* 43(4):2433–2441.
11. Jonas S, Izaurralde E (2013) The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev* 27(24):2628–2641.
12. Protter DSW, Parker R (2016) Principles and Properties of Stress Granules. *Trends Cell Biol* 26(9):668–679.
13. Khong A, et al. (2017) The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol Cell* 0(0). doi:10.1016/j.molcel.2017.10.015.
14. Ribeiro D, et al. (in press) Protein complex scaffolding predicted as a prevalent function of human long non-coding RNAs. *Nucleic Acids Research*. doi:10.1093/nar/gkx1169.
15. Van Nostrand EL, et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13(6):508–514.

16. Berkovits BD, Mayr C (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522(7556):363–367.
17. Saha S, Hyman AA (2017) RNA gets in phase. *J Cell Biol* 216(8):2235–2237.
18. Brouwer JR, Willemsen R, Oostra BA (2009) Microsatellite repeat instability and neurological disease. *BioEssays News Rev Mol Cell Dev Biol* 31(1):71–83.
19. Everett CM, Wood NW (2004) Trinucleotide repeats and neurodegenerative disease. *Brain J Neurol* 127(Pt 11):2385–2405.
20. Todd PK, et al. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78(3):440–455.
21. Strack RL, Disney MD, Jaffrey SR (2013) A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat Methods* 10(12):1219–1224.
22. Tassone F, Iwahashi C, Hagerman PJ (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol* 1(2):103–105.
23. Botta-Orfila T, Tartaglia GG, Michalon A (2016) Molecular Pathophysiology of Fragile X-Associated Tremor/Ataxia Syndrome and Perspectives for Drug Development. *Cerebellum Lond Engl* 15(5):599–610.
24. Iwahashi CK, et al. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain J Neurol* 129(Pt 1):256–271.
25. Sellier C, et al. (2017) Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* 93(2):331–347.
26. Sofola OA, et al. (2007) RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* 55(4):565–571.
27. Sellier C, et al. (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J* 29(7):1248–1261.
28. Sellier C, et al. (2013) Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep* 3(3):869–880.

29. Qurashi A, Li W, Zhou J-Y, Peng J, Jin P (2011) Nuclear Accumulation of Stress Response mRNAs Contributes to the Neurodegeneration Caused by Fragile X Premutation rCGG Repeats. *PLOS Genet* 7(6):e1002102.
30. Marchese D, de Groot NS, Lorenzo Gotor N, Livi CM, Tartaglia GG (2016) Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip Rev RNA* 7(6):793–810.
31. Tartaglia GG (2016) The Grand Challenge of Characterizing Ribonucleoprotein Networks. *Front Mol Biosci* 3. doi:10.3389/fmolb.2016.00024.
32. Disney MD, et al. (2012) A Small Molecule that Targets r(CGG)exp and Improves Defects in Fragile X-Associated Tremor Ataxia Syndrome. *ACS Chem Biol* 7(10):1711–1718.
33. Mittal N, Scherrer T, Gerber AP, Janga SC (2011) Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. *J Mol Biol* 409(3):466–479.
34. Huttlin EL, et al. (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162(2):425–440.
35. Armaos A, Cirillo D, Tartaglia GG (2017) omiXcore: a web server for prediction of protein interactions with large RNA. *Bioinforma Oxf Engl*. doi:10.1093/bioinformatics/btx361.
36. Clemson CM, et al. (2009) An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell* 33(6):717–726.
37. Lee S, et al. (2016) Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* 164(1–2):69–80.
38. Tichon A, et al. (2016) A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat Commun* 7:ncomms12209.
39. Zhang H, et al. (2015) RNA controls PolyQ protein phase transitions. *Mol Cell* 60(2):220–230.
40. Reineke LC, Kedersha N, Langereis MA, Kuppeveld FJM van, Lloyd RE (2015) Stress Granules Regulate Double-Stranded RNA-Dependent Protein Kinase Activation through a Complex Containing G3BP1 and Caprin1. *mBio* 6(2):e02486-14.
41. Jain A, Vale RD (2017) RNA phase transitions in repeat expansion disorders. *Nature* 546(7657):243–247.
42. Krzyzosiak WJ, et al. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res* 40(1):11–26.

43. Delli Ponti R, Marti S, Armaos A, Tartaglia GG (2017) A high-throughput approach to profile RNA structure. *Nucleic Acids Res* 45(5):e35–e35.
44. Agostini F, et al. (2013) catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinforma Oxf Engl* 29(22):2928–2930.
45. Livi CM, Klus P, Delli Ponti R, Tartaglia GG (2015) catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinforma Oxf Engl*. doi:10.1093/bioinformatics/btv629.
46. Marchese D, et al. Discovering the 3′ UTR-mediated regulation of alpha-synuclein. *Nucleic Acids Res*. doi:10.1093/nar/gkx1048.
47. Cirillo D, et al. (2017) Quantitative predictions of protein interactions with long noncoding RNAs. *Nat Methods* 14(1):5–6.
48. Cirillo D, et al. (2013) Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. *RNA N Y N* 19(2):129–140.
49. Warde-Farley D, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 38(Web Server issue):W214–W220.
50. Klus P, Ponti RD, Livi CM, Tartaglia GG (2015) Protein aggregation, structural disorder and RNA-binding ability: a new approach for physico-chemical and gene ontology classification of multiple datasets. *BMC Genomics* 16(1):1071.
51. Morris MJ, Wingate KL, Silwal J, Leeper TC, Basu S (2012) The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res* 40(9):4137–4145.
52. Budny B, et al. (2010) Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. *Clin Genet* 77(6):541–551.
53. Procaccio V, et al. (2006) A mutation of beta -actin that alters depolymerization dynamics is associated with autosomal dominant developmental malformations, deafness, and dystonia. *Am J Hum Genet* 78(6):947–960.
54. Rivière J-B, et al. (2012) De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nat Genet* 44(4):440–444, S1-2.

55. Narkis G, et al. (2007) Lethal contractural syndrome type 3 (LCCS3) is caused by a mutation in PIP5K1C, which encodes PIPKI gamma of the phophatidylinsitol pathway. *Am J Hum Genet* 81(3):530–539.
56. Erdmann J, et al. (2003) Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. *Clin Genet* 64(4):339–349.
57. de Silva MG, et al. (2003) Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an attention deficit hyperactivity disorder-like phenotype. *J Med Genet* 40(10):733–740.
58. Wood JD, et al. (1998) Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. *Mol Cell Neurosci* 11(3):149–160.
59. Zou W, et al. (2011) The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. *Nat Cell Biol* 13(1):59–65.
60. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 42(D1):D92–D97.
61. Maurin T, Zongaro S, Bardoni B (2014) Fragile X Syndrome: from molecular pathology to therapy. *Neurosci Biobehav Rev* 46 Pt 2:242–255.
62. Buijsen RAM, et al. (2014) FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun* 2:162.
63. Hukema RK, et al. (2015) Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum Mol Genet* 24(17):4948–4957.
64. Banani SF, Lee HO, Hyman AA, Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18(5):285–298.
65. Markmiller S, et al. (2018) Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* 172(3):590-604.e13.
66. Tacke R, Tohyama M, Ogawa S, Manley JL (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93(1):139–148.
67. Haynes C, Iakoucheva LM (2006) Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res* 34(1):305–312.

68. Ying Y, et al. (2017) Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* 170(2):312-323.e10.
69. Gueroussov S, et al. (2017) Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* 170(2):324-339.e23.
70. Zlotorynski E (2017) Splicing: Phasing alternative exons. *Nat Rev Mol Cell Biol* 18(9):529.
71. Tyson-Capper A, et al. (2014) Human Tra2 proteins jointly control a *CHEK1* splicing switch among alternative and constitutive target exons. *Nat Commun* 5:ncomms5760.
72. Bardoni B, Davidovic L, Bensaid M, Khandjian EW (2006) The fragile X syndrome: exploring its molecular basis and seeking a treatment. *Expert Rev Mol Med* 8(8):1–16.
73. Hlevnjak M, Polyansky AA, Zagrovic B (2012) Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. *Nucleic Acids Res* 40(18):8874–8882.
74. Zanzoni A, et al. (2013) Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. *Nucleic Acids Res* 41(22):9987–9998.
75. Knowles TPJ, et al. (2009) An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* 326(5959):1533–1537.
76. Tartaglia GG, Vendruscolo M (2009) Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. *Mol Biosyst* 5(12):1873–1876.
77. Speretta E, et al. (2012) Expression in Drosophila of Tandem Amyloid β Peptides Provides Insights into Links between Aggregation and Neurotoxicity. *J Biol Chem* 287(24):20748–20754.
78. Bellucci M, Agostini F, Masin M, Tartaglia GG (2011) Predicting protein associations with long noncoding RNAs. *Nat Methods* 8(6):444–445.
79. Mateu-Huertas E, et al. (2014) Blood expression profiles of fragile X premutation carriers identify candidate genes involved in neurodegenerative and infertility phenotypes. *Neurobiol Dis* 65:43–54.

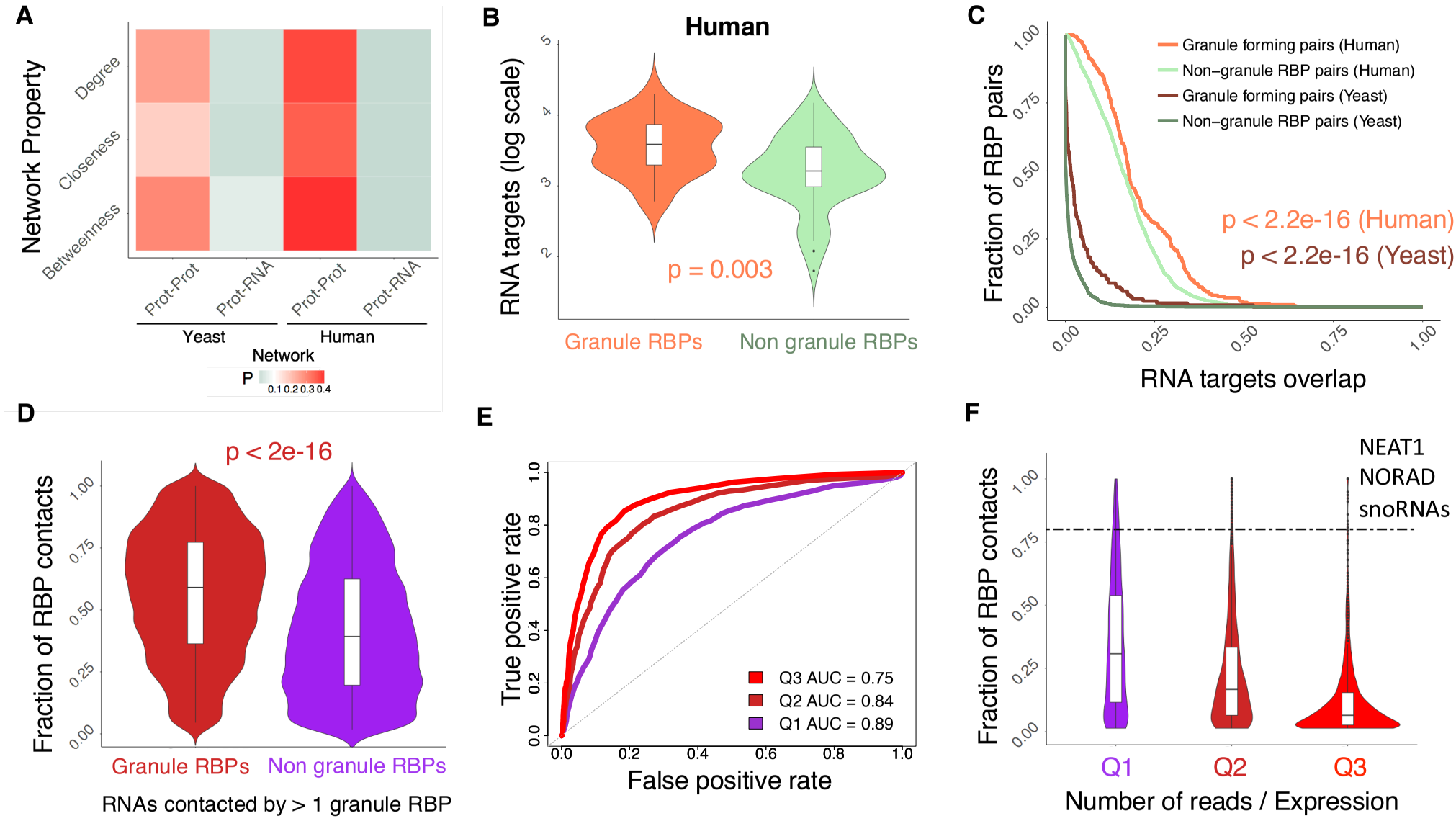

A
Network Property
Degree
Closeness
Betweenness
Prot-Prot
Prot-RNA
Prot-Prot
Prot-RNA
Yeast
Human
Network
P
0.1 0.2 0.3 0.4
B
Human
RNA targets (log scale)
p = 0.003
Granule RBPs
Non granule RBPs
C
Fraction of RBP pairs
Granule forming pairs (Human)
Non–granule RBP pairs (Human)
Granule forming pairs (Yeast)
Non–granule RBP pairs (Yeast)
p < 2.2e-16 (Human)
p < 2.2e-16 (Yeast)
RNA targets overlap
D
p < 2e-16
Fraction of RBP contacts
Granule RBPs
Non granule RBPs
RNAs contacted by > 1 granule RBP
E
True positive rate
False positive rate
Q3 AUC = 0.75
Q2 AUC = 0.84
Q1 AUC = 0.89
F
NEAT1
NORAD
snoRNAs
Fraction of RBP contacts
Q1
Q2
Q3
Number of reads / Expression

Figure 1

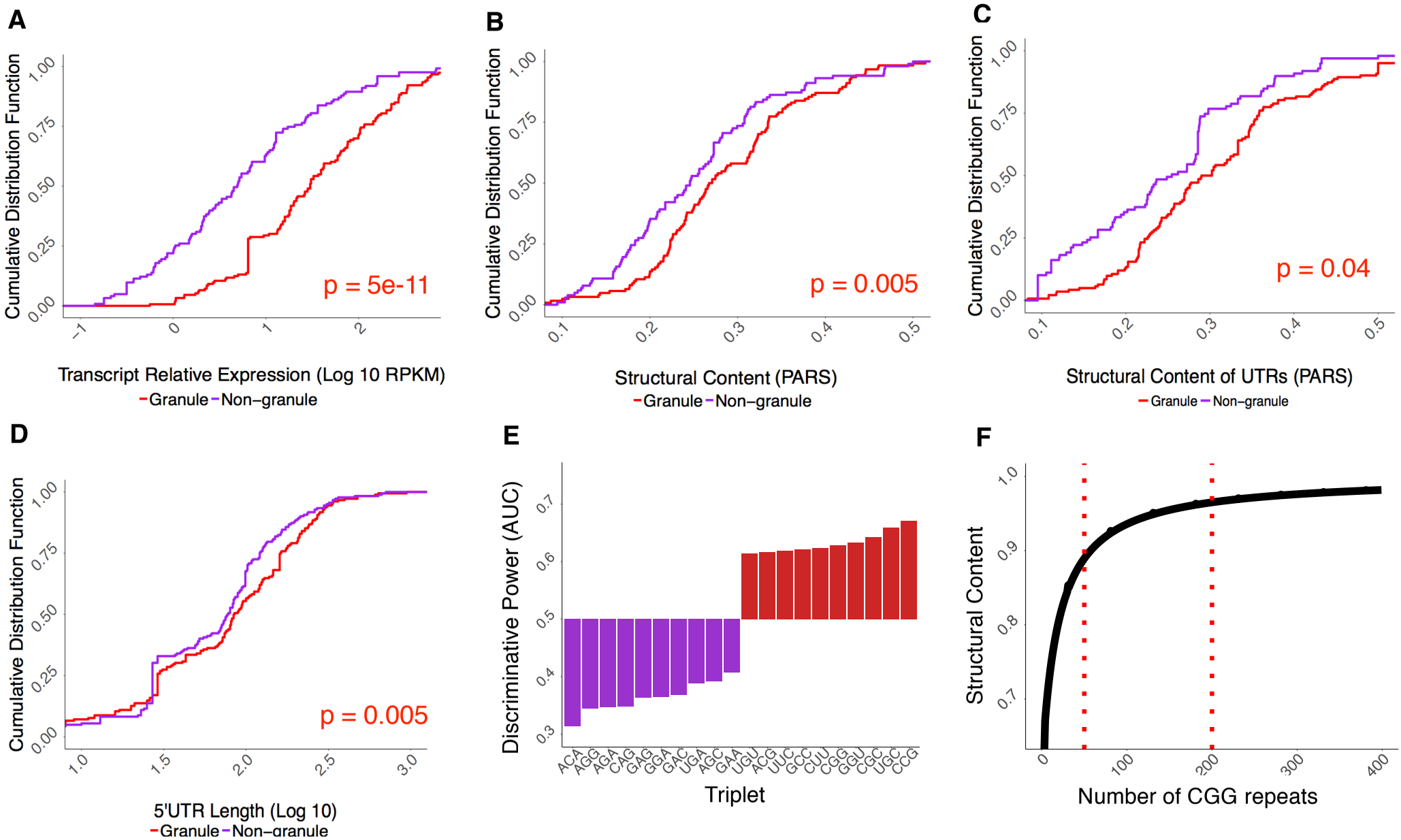
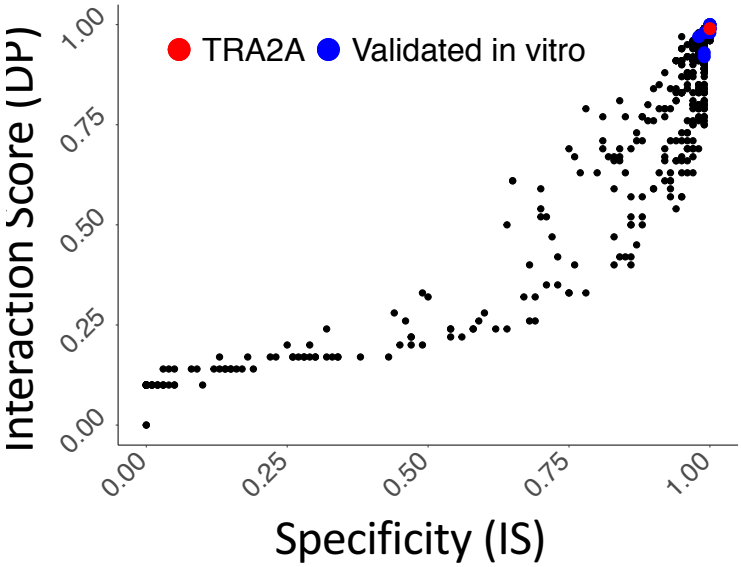
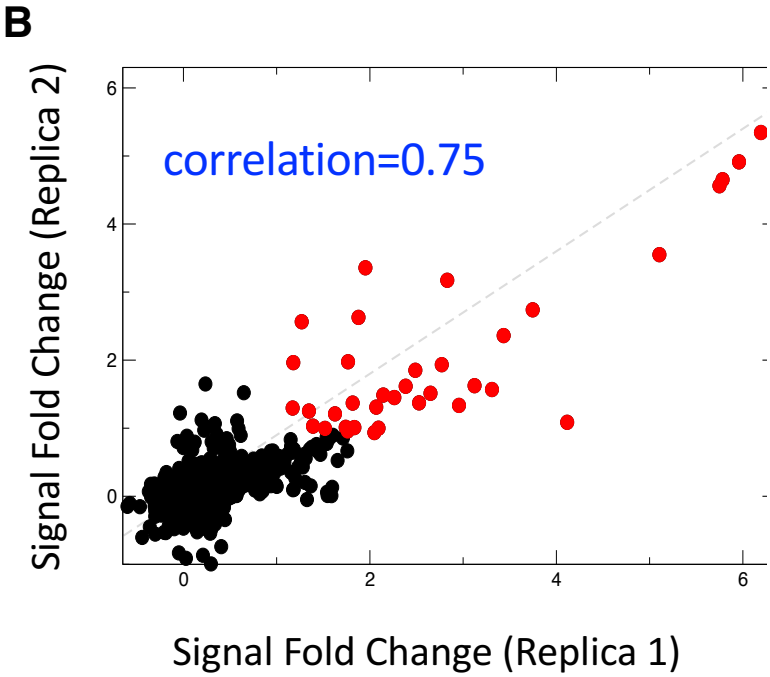
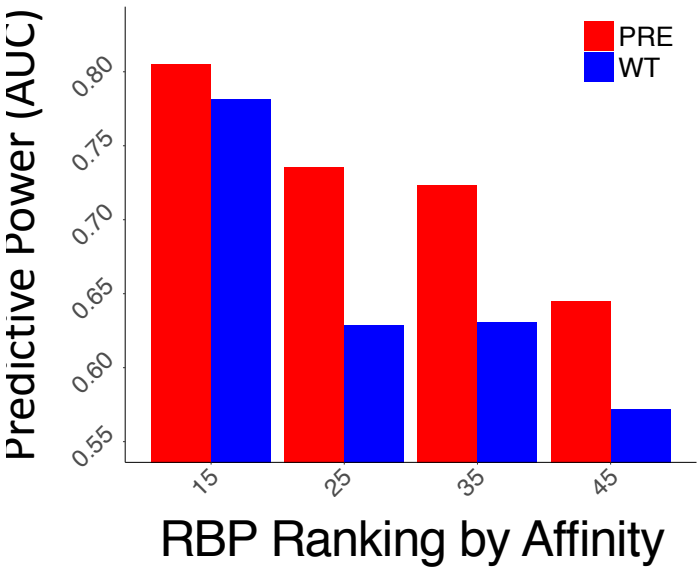

A
Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
−1
0
1
2
p = 5e-11
Transcript Relative Expression (Log 10 RPKM)
Granule
Non−granule
B
Cumulative Distribution Function
0.1
0.2
0.3
0.4
0.5
p = 0.005
Structural Content (PARS)
Granule
Non−granule
C
Cumulative Distribution Function
p = 0.04
Structural Content of UTRs (PARS)
Granule
Non−granule
D
Cumulative Distribution Function
1.0
1.5
2.0
2.5
3.0
p = 0.005
5'UTR Length (Log 10)
Granule
Non−granule
E
Discriminative Power (AUC)
0.3
0.4
0.5
0.6
0.7
ACA
AGG
AGA
CAG
GAG
GGA
GAC
UGA
AGC
GAA
UGU
ACG
UUC
GCC
CUU
CGG
GGU
CGC
UGC
CCG
Triplet
F
Structural Content
0.7
0.8
0.9
1.0
0
100
200
300
400
Number of CGG repeats

Figure 2


TRA2A
Validated in vitro
Interaction Score (DP)
Specificity (IS)
0.00
0.25
0.50
0.75
1.00


B
correlation=0.75
Signal Fold Change (Replica 2)
Signal Fold Change (Replica 1)
0
2
4
6


PRE
WT
Predictive Power (AUC)
RBP Ranking by Affinity
0.55
0.60
0.65
0.70
0.75
0.80
15
25
35
45

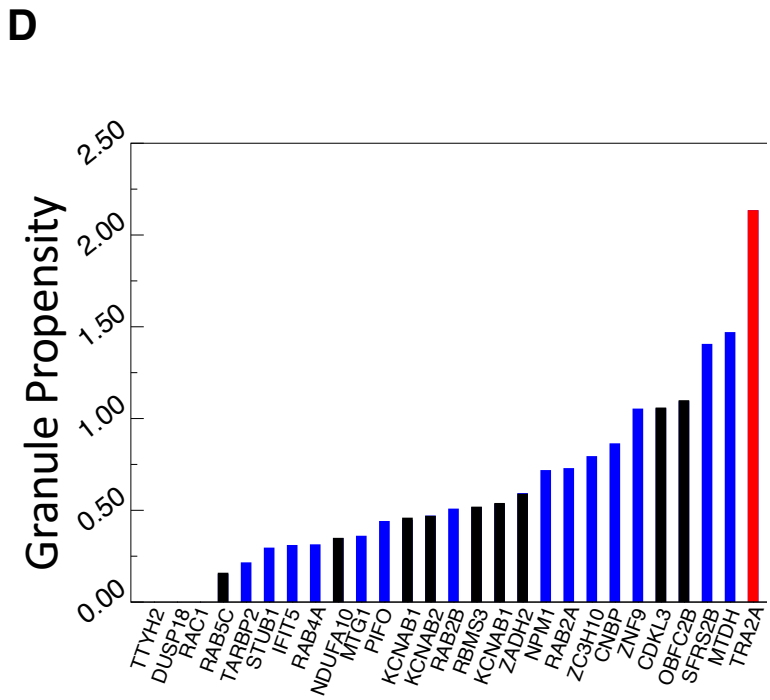

D
Granule Propensity
0.00
0.50
1.00
1.50
2.00
2.50
TTYH2
DUSP18
RAC1
RAB5C
TARBP2
STUB1
IFIT5
RAB4A
NDUFA10
MTG1
PIFO
KCNAB1
KCNAB2
RAB2B
RBMS3
KCNAB1
ZADH2
NPM1
RAB2A
ZC3H10
CNBP
ZNF9
CDKL3
OBFC2B
SFRS2B
MTDH
TRA2A

Figure 3

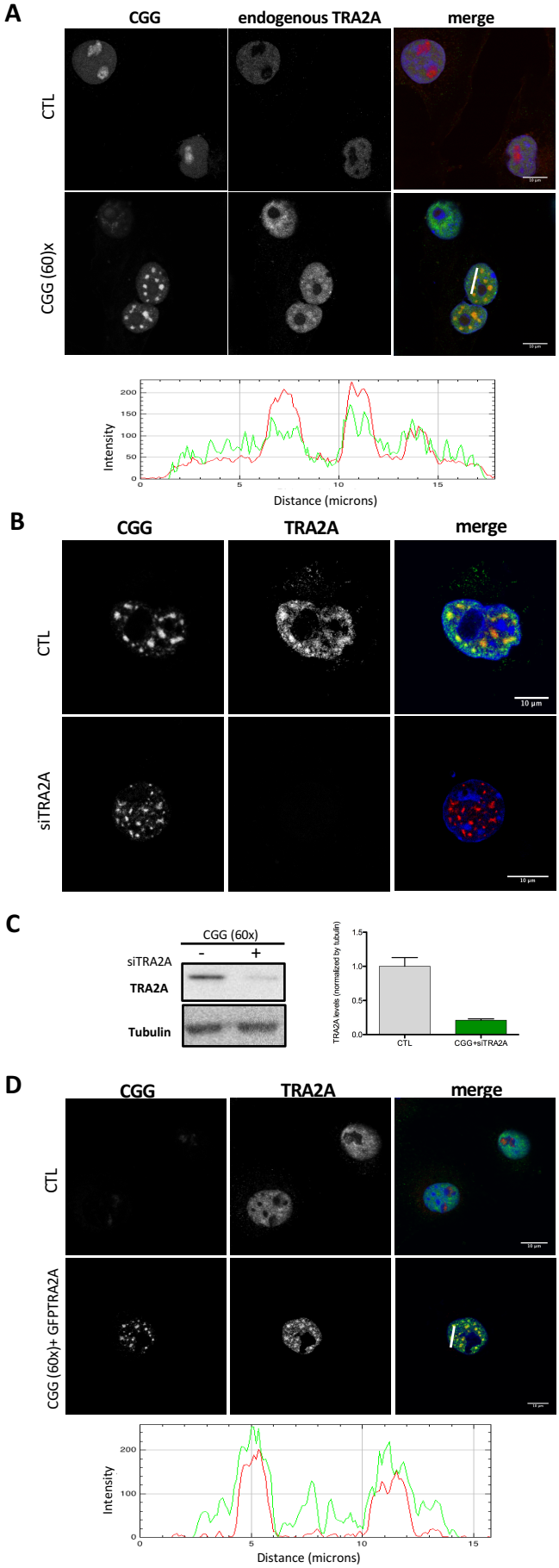

A
CGG
endogenous TRA2A
merge
CTL
CGG (60)x
Intensity
Distance (microns)
B
CGG
TRA2A
merge
CTL
siTRA2A
C
CGG (60x)
siTRA2A
-
+
TRA2A
Tubulin
TRA2A levels (normalized by tubulin)
CTL
CGG+siTRA2A
D
CGG
TRA2A
merge
CTL
CGG (60x)+ GFPTRA2A
Intensity
Distance (microns)

Figure 4

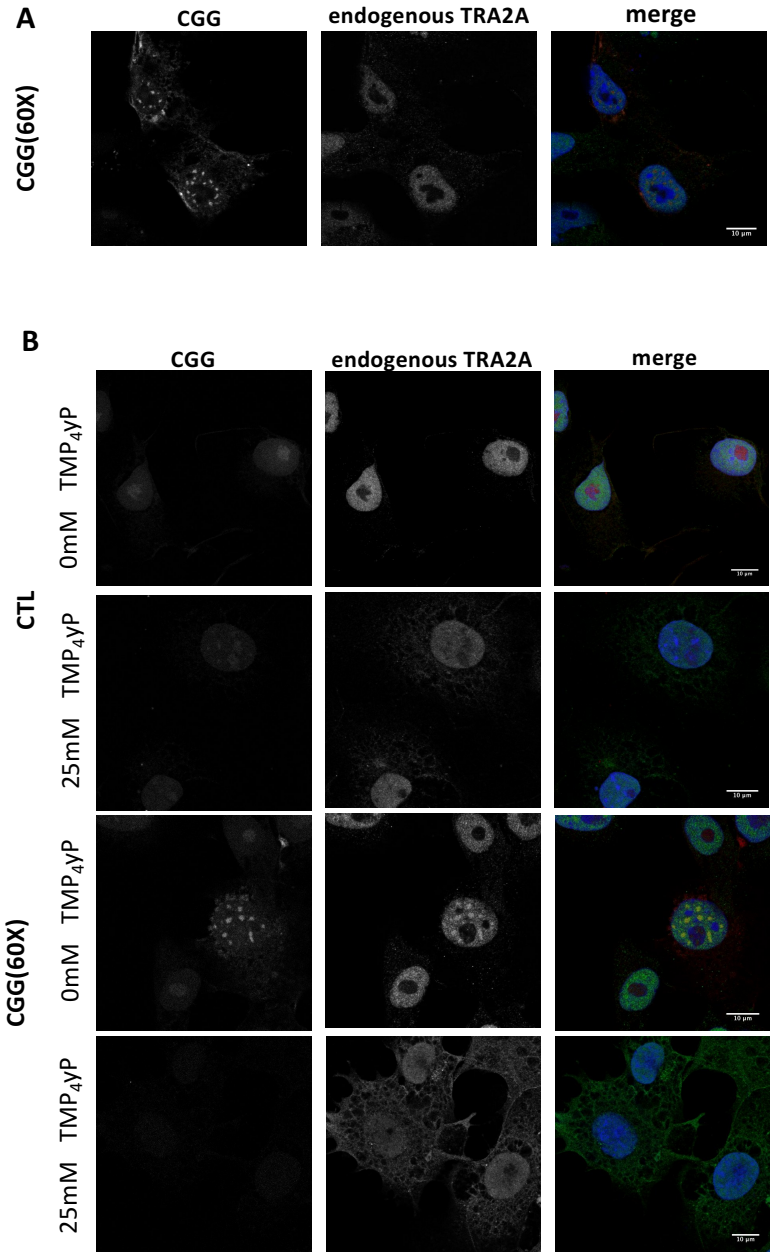

A
CGG
endogenous TRA2A
merge
CGG(60X)
B
CGG
endogenous TRA2A
merge
CTL
0mM TMP4yP
25mM TMP4yP
CGG(60X)
0mM TMP4yP
25mM TMP4yP

Figure 5

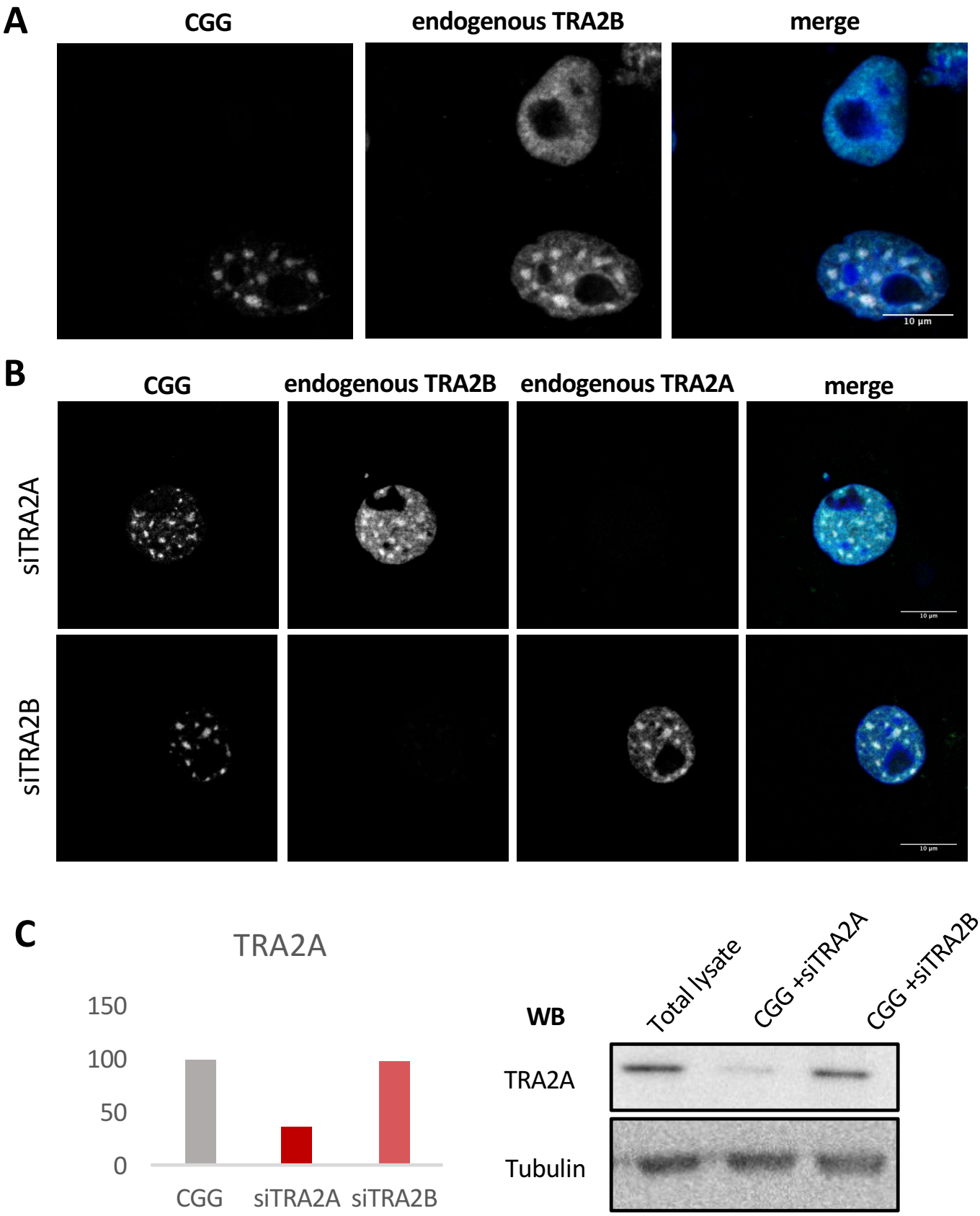

A
CGG
endogenous TRA2B
merge
10 µm
B
CGG
endogenous TRA2B
endogenous TRA2A
merge
siTRA2A
siTRA2B
C
TRA2A
150
100
50
0
CGG
siTRA2A
siTRA2B
WB
Total lysate
CGG +siTRA2A
CGG +siTRA2B
TRA2A
Tubulin

Figure 6

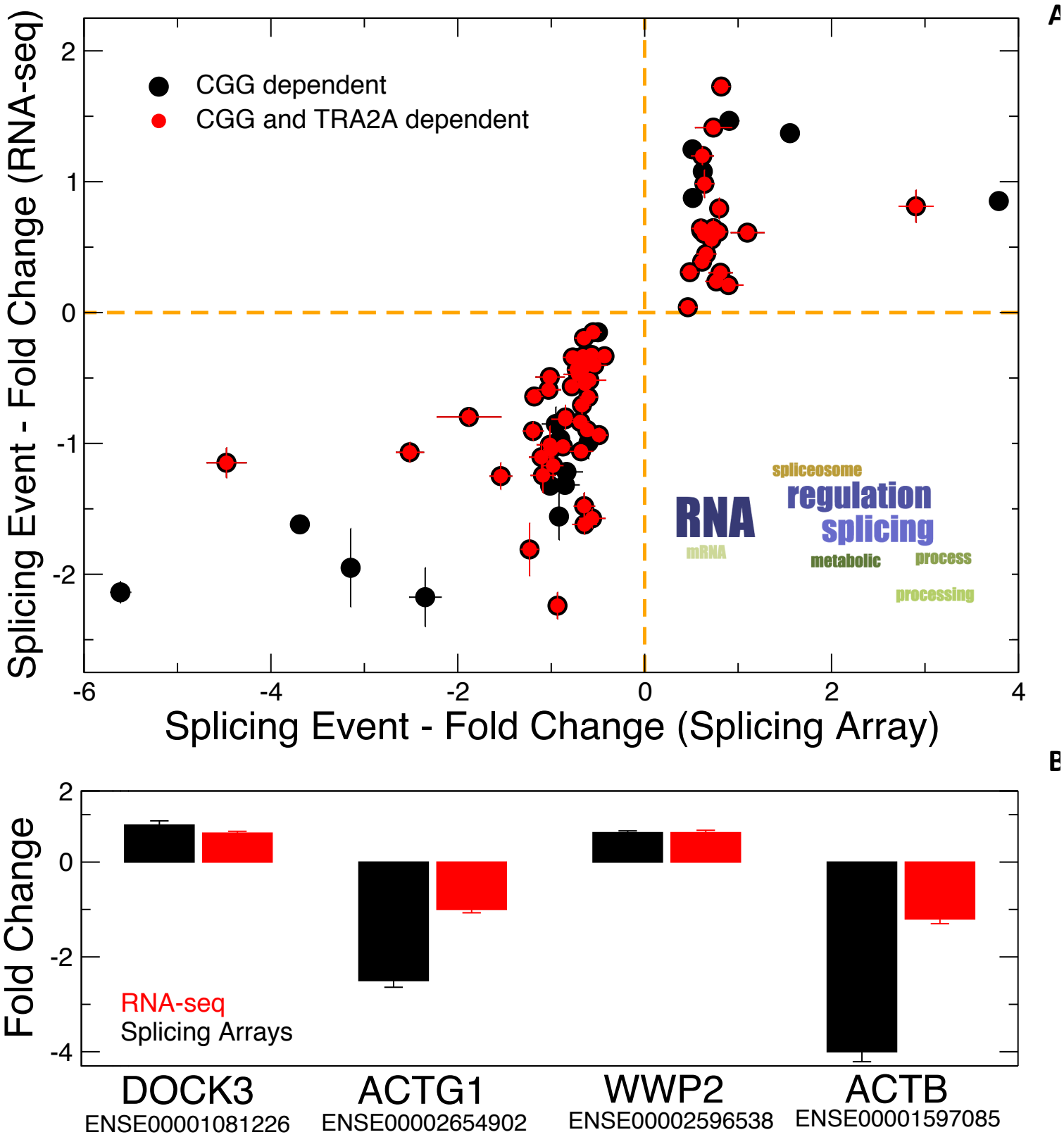

A
CGG dependent
CGG and TRA2A dependent
Splicing Event - Fold Change (RNA-seq)
Splicing Event - Fold Change (Splicing Array)
2
1
0
-1
-2
-6
-4
-2
0
2
4
RNA
mRNA
spliceosome
regulation
splicing
metabolic
process
processing
B
Fold Change
2
0
-2
-4
RNA-seq
Splicing Arrays
DOCK3
ENSE00001081226
ACTG1
ENSE00002654902
WWP2
ENSE00002596538
ACTB
ENSE00001597085

Figure 7

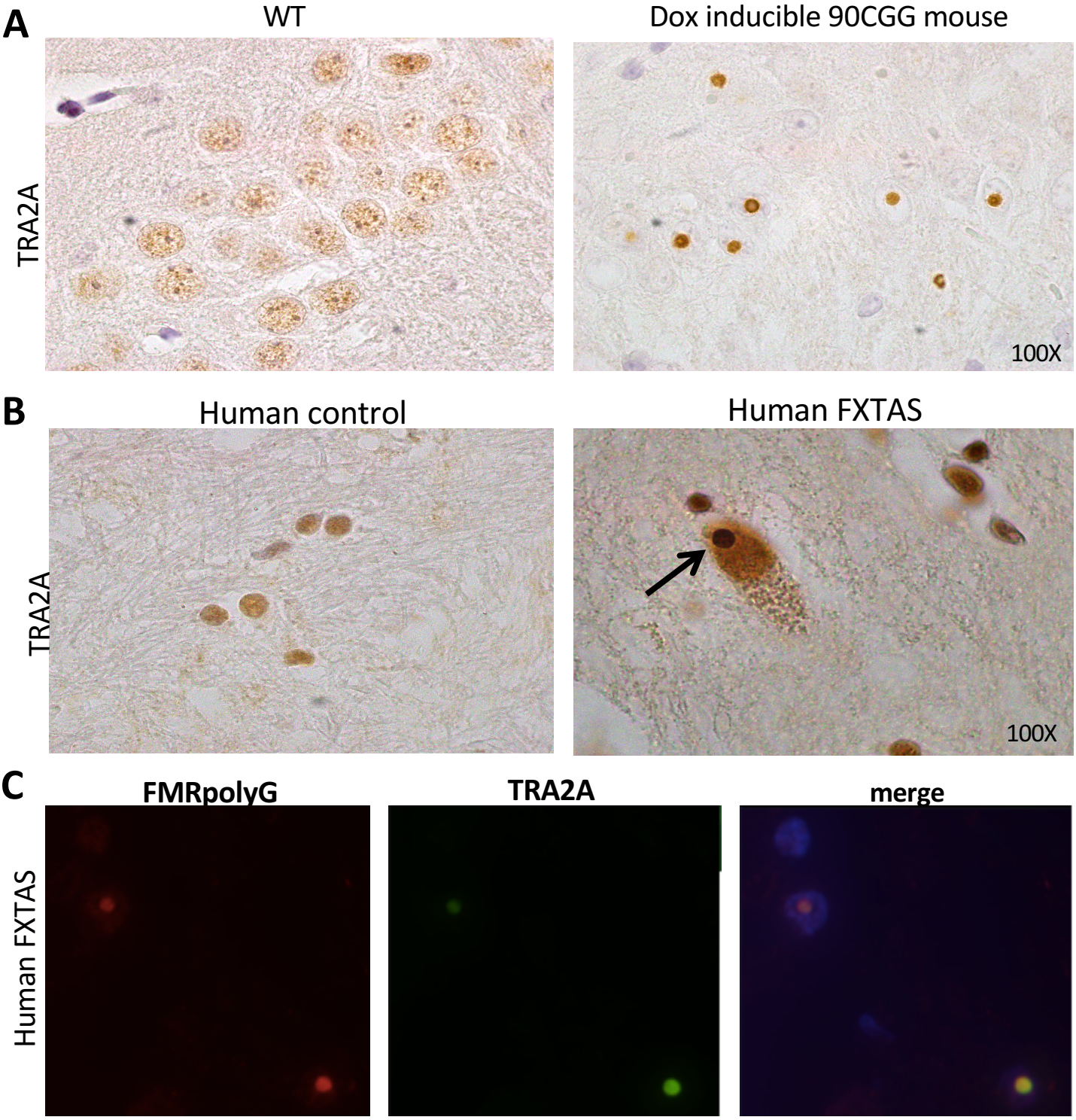

A
WT
Dox inducible 90CGG mouse
TRA2A
100X
B
Human control
Human FXTAS
TRA2A
100X
C
FMRpolyG
TRA2A
merge
Human FXTAS

Figure 8

Computational analysis

Granule-forming proteins were extracted from a previous publication (1) that reports the most exhaustive list of components for cytoplasmic RNP granules to date, comprising 205 yeast and 411 human granule-forming proteins (**Supplementary Figure 1A; Supplementary Table 1A**).

Human protein-protein interactions were taken from the BioPlex database (2) that includes highly curated data produced by high-throughput affinity-purification mass spectrometry. BioPlex contains 56,554 interactions among 510,882 different proteins **(Figure 1A; Supplementary Figure 2B)**. Human protein-RNA interactions were identified through eCLIP experiments (3). The dataset contains 1,103,800 interactions of 78 proteins in the K562 cell line **(Figure 1B; Supplementary Figure 2B)**. We processed the eCLIP data normalizing the number of reads by gene expression (4). We considered interactions having values of number of reads by expression level higher than the first, second and third quartile of the distribution of the whole dataset **(Figure 1F**; **Supplementary Figure 3; Supplementary Table 1B, 1C** and **1D**). After data normalization and filtering, the dataset includes 22,961 transcripts interacting with at least one protein (20,724 coding and 2,237 non-coding). We extracted the expression levels for K562 transcripts from the ENCODE project (4).

For yeast analysis we used the dataset reported in a previous article (5) that includes both protein-protein and protein-RNA interactions. The protein-protein network is based on the integration of two mass spectrometry studies that comprise a total of 5,303 proteins and 401,821 interactions (5) **(Figure 1A)**. Protein-RNA interactions were extracted from the integration of four different studies on immunoprecipitation of RBPs followed by microarray analysis of the bound transcripts (**Supplementary Table 1E**). The data includes a total of 24,932 interactions from 69 RPBs to 6,159 transcripts.

We considered non-granule forming those RBPs present in the protein-RNA dataset and not described as granule-forming in a previous article (1) (**Supplementary Figure 1; Supplementary Table 1B, 1C** and **1D**). There are 22,571 transcripts interacting with at least one granule-forming protein, 287 transcripts interacting only with granule-forming proteins and 390 with only non-granule forming proteins.

In the case of the granule and non-granule protein-protein networks comparison, we included RBP lists provided by (Brannan et al., 2016;Gerstberger et al., 2014) for yeast and human. These datasets comprise a total of 690 yeast RBPs and 1,795 human RBPs.

Network analysis

Protein-protein and protein-RNA networks consisted of a set of nodes (protein or RNAs) that are connected through edges (interactions). All network analyses were performed in the R environment (http://www.r-project.org) using the igraph package (http://igraph.org/)(Nepusz T). We employed build-in functions to compute degree, betweenness and closenesss measures of centrality. Networks were considered directionless and unweighted. Degree centrality is defined as the number of edges a node has. The other centrality measures were based on the shortest path length between nodes in the network (i.e. minimum number of edges between two certain nodes). In this sense, betweenness is defined as the number of shortest paths in the

network that go through a certain node. Closessness centrality is the inverse of the average of the shortest path between a certain node and all the other nodes in the network. We compared the distribution of centrality values for granule and non-granule RBPs in the same global protein-protein or protein-RNA network **(Figure 1A; Supplementary Figure 2).**

We used the Jaccard index (**Supplementary Tables 1F, 1G, 1H** and **1I**) as a measure of the overlap of RNA targets between pairs of proteins. The Jaccard index of a specific couple of proteins a and b ($J_{a,b}$) was computed as:

$$J_{a,b} = \frac{|A \cap B|}{|A \cup B|}$$

were A is the set of RNA targets of the first protein of the pair, B is the set of RNA targets of the second protein of the pair, $|A \cap B|$ is the size of the intersection of A and B (i.e. number of RNA targets shared by the two proteins) and $|A \cup B|$ is the size of the union of A and B (i.e. the total number of RNA targets of A and B minus the number of shared RNA targets).

RNA properties analysis

To study features of granule and non-granule transcripts, we compared RNAs with the same number of total proteins contacts (**Supplementary Table 1A**). We used transcripts with a total of 10 RBP contacts in the human analysis and 5 contacts in the yeast case. The number of RBP contacts was chosen to have granule and non-granule sets of comparable size (**Supplementary Tables 2B** and **2C**). Granule transcripts are contacted by a larger number of granule-forming RBPs than non-granule forming RBPs (*viceversa* for non-granule transcripts).

The UTR analysis is based on Ensembl annotation (**Supplementary Tables 2B** and **2C**).

RNA secondary structure

To profile the secondary structure of granule and non-granule transcripts **(Figure 2E; Supplementary Figures 2B, 4C** and **4D; Supplementary Tables 2B** and **2C**), we used PARS (Parallel Analysis of RNA Structure) data (9, 10). PARS distinguishes double- and single-stranded regions using the catalytic activity of two enzymes, RNase V1 (able to cut double-stranded nucleotides) and S1 (able to cut single-stranded nucleotides). Nucleotides with a PARS score higher than 0 indicate double-stranded conformation, while values lower than 0 are considered single-stranded (9, 10). Undetermined nucleotides with a PARS score of 0 were discarded from our analysis. For each transcript, we counted the number of nucleotides with PARS score above zero divided by total length of the sequence.

We also predicted the secondary structure of granule and non-granule transcripts using CROSS (PARS human model). Input sequences were the same employed for the granule RNA properties analysis. For each sequence, structural content is defined as the percentages of nucleotides with a CROSS score bigger than 0.5 (double-stranded prone).

Statistical analysis

To assess whether granule RBPs exhibit different trends compared to non-granule RBPs, we

used the Wilcoxon test (also called Mann-Whitney U test). Wilcoxon test is a non-parametric test used to compare the mean of two distributions without any given assumption about them. To compare properties of highly versus lowly contacted RNAs and difference in target overlap between granule and non-granule pairs, we used the Kolmogorov-Smirnov test (KS test). KS test is also a non-parametric test used to compare the distance between two cumulative distribution functions (CDFs).

catRAPID omics analysis

*cat*RAPID *omics* was used to compute the interaction propensity of CGG repeats with proteins (11). *cat*RAPID *omics* ranks predictions based on interactions score as well as presence of motifs and RBDs (12). For RNA sequences > 1000 nt, the uniform fragmentation procedure is applied to determine the binding regions of a protein.

catGRANULE analysis

Structural disorder, nucleic acid binding propensity and amino acid patterns such as arginine-glycine and phenylalanine-glycine are key features of proteins coalescing in granules(13). These features were combined in a computational approach, *cat*GRANULE, that we employed to identify RBPs assembling into granules (scores > 0 indicate granule propensity)(13).

Experimental details

RNA IVT and Protein arrays

FMR1 5'-UTR expanded and control pCAGIG vectors, with 79 and 21 repeats, respectively, were generated by Dr. Marti's group (14). The UTRs were subcloned in PBSK plasmid containing promoters suitable for in vitro transcription. The plasmid was digested in final volume reaction of 30ul with restriction enzymes and the digestion was ensured by loading 1ul in a 1% agarose gel. The reaction was purified with the MinElute PCR Purification Kit following manufacturer's instructions. In vitro transcription was performed with the T7 Megascript T7, High Yield Transcription Kit, Invitrogen, Thermo Scientific according to standard procedure with the addition of 1% DMSO and 1% ribolock, overnight at 37°C. The synthetized RNA was treated with TURBODNAse 2U/ul (Invitrogen) at 37°C for 15min. The RNA was purified with magnetic beads (Agencourt RNA Clean XP) eluting in 30ul of nuclease-free water.
The integrity and specificity of the RNA was checked by means of RNA denaturing agarose gel and Bioanalyzer quality control.

The CGGxRNA was fluorescently labeled with Cy5 Label IT uArray Labeling Kit (Mirus) with slight modifications from standard protocol. Briefly, 5ug of RNA were mixed with 1:5 Label IT Cy5 reagent and incubated in a final volume of 25ul at 37°C for 70min. The reaction was stopped by adding 2.5ul of 10X Stop solution. Again the labeled RNA was purified with magnetic beads (Agencourt RNA Clean XP).

The RNA concentration and labeling density were measured with Nanodrop 1000 spectrophotometer (Thermo Scientific) and calculated as follows.

Only reactions with an RNA labeling density of 1 Cy5 dye per 700-900 nt were used.

Base:dye = (Abase*ε dye)/(Adye*ε base)
Abase=A260-(Adye*CF260)

Constants:
ε dye= 250000
CF260= 0.05
ε base= 8250

Labeled RNA integrity was verified with the Agilent 2100 Bioanalyzer.

50 pmoles of labeled RNA were hybridized in the protein arrays Human Protein Microarrays v5,2, Life Technologies.

The arrays were dried and immediately scanned at 635nm in Microarray Scanner G2505B (Agilent). GenePix Pro 6.1 software (Molecular Devices) was used to determine the signal at 635nm of each spotted protein location and therefore quantify the RNA-protein interaction. Specifically, the local background intensity (B635) was subtracted from the intensity (F635) at each of the duplicate spots for a given protein, to quantify. Data was filtered based on signal to background ratio for each of the duplicate feature to be greater than 2.5 fold and Z-Score ≥ 3 from the global mean signal from all the spotted proteins. Finally, the intersection of technical replicates was considered as the final value for quantification.

Cell Culture

Human lymphocytes cells from Coriell repository (CGG(41X); Coriell repository number NA20244A and CGG(90X) Coriell repository number GM06906B) were grown in suspension in DMEM 10% fetal bovine serum (FBS) 1% Penicillin/Streptomycin, 2mM Glutamine at 37ºC with a 5% CO2 atmosphere. Cell counting was performed with Neubauer chamber.

COS-7 cell lines were cultured in DMEM 10% FBS, 0,1% non-essential aminoacids, pyruvate and glutamine, at 37ºC with a 5% CO2 atmosphere. Cells were counted with a Neubauer chamber.

For human fibroblast culture, skin biopsies were obtained using a 3-mm punch from 3 patients with FXTAS and 3 control individuals recruited from the Dermatology Department of the Hospital Clínic of Barcelona with the corresponding informed consent. The biopsy was diced under sterile conditions and then plated at 37ºC with a 5% CO2 atmosphere in T25 flasks in MEM 13% Newborn Calf Serum, 0.06% Penicillin and 0.06% Streptomycin. Trypan blue exclusion test was used to quantify the number cell viability and cell counts using a haemocytometer.

IF-RNA FISH in COS-7 cells

COS-7 cells were grown on 13mm coverslips until a 70% confluence. Cells were transfected with lipofectamine 2000 (Invitrogen, #13778150) according to manufacturer's instructions and stained after 24 hours.

SiRNA treatments were performed with lipofectamine 2000 (Invitrogen, 11668019) transfection of (siTRA2A, Ambion AM16704, siTRA2B, Ambion S12749, compared to water). Overexpression of proteins was achieved by transfection of GFP-TRA2A, GFP-TRA2B, compared with GFP-vector, all plasmids kindly given by Dr. Nicolas Charlet-Berguerand. For further treatments, cells were incubated with 25nM TMP4yP (Abcam ab120793) or 15uM of 1a molecule (this latter kindly given by Matthew D. Disney).

Prior to immunostaining, cells were fixed with 4% paraformaldehyde for 10 minutes and washed three times with PBS. Permeabilization was done with Triton X-100 0.1% for 5 minutes. Cells were washed 3 times with PBS and then blocked with BSA 1% solution for 20 minutes and washed again with PBS. Primary antibodies were used in a 1:50 dilution (antiTRA2A, Abcam, ab72625, antiTRA2B, Abcam, ab31353). Secondary antibodies (anti Rabbit 647, ab150115, anti-rabbit 488, ab11008) were used in a 1:200 dilution after three washes with PBS solution. RNA FISH assay was done after the immunostaining according to manufacturer's protocol (Stellaris, Biosearch Technologies). The RNA FISH probe (GGC)8X-Cy3, Sigma) was used at 125nM final concentration and cells were then incubated at 37ºC overnight, as reported in Sellier et al 2010. Finally, cells were mounted directly in Fluoroshield with DAPI histology mounting medium (Sigma, #F6057). All coverslips were examined using a fluorescence microscope (Leica) coupled to a DMI600 camera. Intensity graphs were generated with Image J software to assess levels of colocalization of signals from different fluorescence channels.

q-PCR

Human lymphocytes were washed and pelleted by centrifugation at 800rpm for 2 min. COS-7 cells and fibroblasts were tripsinized and pelleted by centrifugation at 1200rpm for 2 minutes. RNA extraction from the different cultured cells was done according to manufacturer's instructions (Qiagen, #50974136). cDNA was generated by RT-PCR using SuperScript III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, # 11752250) to quantify mRNAs. q-PCR was performed using Sybr Green master mix (Invitrogen, # 4367659) and analyzed by AB7900HT (Leica). In all experiments, GAPDH was used as internal control in all experiments.

Western Blot

Total proteins from human lymphocytes, fibroblasts and COS-7 cells (one day post-transfection) were extracted. The level of protein was measured by the Bio-Rad Protein Assay according to manufacturer's instructions. All lysates were resolved in a 4-12% gel (NuPAGE, Invitrogen) according to the molecular size of the proteins–and then transferred to a nitrocellulose membrane 0.2µ m. The membranes were blocked with 5% non-fat dry milk in TBS-Tween 1% then washed with PBS and incubated with anti TRA2A (1:1000, Abcam ab72625), anti TRA2B (1:500, Abcam ab31353) or anti Tubulin (1:5000, Abcam ab7291)

overnight at 4ºC. After primary antibodies treatment, membranes were washed three times with TBS-Tween 1% and then incubated with the secondary peroxidase antibody 1h with an anti-mouse (Abcam ab97046) or an anti-rabbit antibody (Protein G, Life Technologies #18-161). Visualization of the signal was achieved by Luminata Starter kit (Millipore, WBLUM0100) according to manufacturer's recommended instructions, and with Amersham Imager 600.

Immunohistochemistry and immunofluorescence from murine and human brain tissue

Tissues were fixed overnight in 4% paraformaldehyde and embedded in paraffin according to standard protocols. Sections (6µm) were deparaffinized followed by antigen retrieval using microwave treatment in 0.01M sodium citrate. Endogenous peroxidase activity was blocked and immunostaining was performed overnight at 4°C using TR2A (Abcam, ab72625) and 8FM 1:10 antibodies(15). In order to better visualize inclusions an extra antigen retrieval step was added, using proteinase K. Antigen-antibody complexes were visualized by incubation with DAB substrate (Dako, K3468) after incubation with Brightvision poly-HRP-linker (Immunologic, DPVO-HRP 55). Slides were counterstained with haematoxylin and mounted with Entellan.

For (double) immunofluorescence, slides were blocked for auto-fluorescence with Sudan Black in 70% ethanol. Primary antibodies include TR2A (Abcam, ab72625), 8FM 1:10(15) and ubiquitin (Dako, Z0458). Secondary antibodies were antirabbit Fab 488 (Molecular Probes, A11070) and antimouse cy3 (Jackson, 715-165-150). Nuclei were visualized with Hoechst.

Splicing Arrays experiments and analysis

COS-7 cells were grown in P10 plates and cultured in different conditions and in three biological replicates each: control, (CGG)60X 185ng, siTRA2A 50nM, (CGG)60X 185ng+siTRA2A 13.6ng, GFP-TRA2A 200ng, (CGG)60X 185ng + GFP-TRA2A 200ng.

Total RNA extraction was performed with Qiagen RNeasy Mini Kit including DNAse treatment according to manufacturer's instructions. RNA amount was quantified and controlled with Nanodrop and Bioanalyzer. 100ng of total RNA from each sample were labelled according to the Affymetrix GeneChip® Whole Transcript Plus protocol, and hybridized to Affymetrix Human Clariom D array using a Affymetrix GeneChip Hybridization Oven 645, in Servei de Microarrays (IMIM-Barcelona). GeneChip was scanned using Affymetrix GeneChip Scanner 3000 7G. The data were analyzed using the RMA algorithm and then LIMMA was applied to calculate significant differential expression between samples. Splicing arrays were analyzed with the Transcriptome Analysis Console Software (Thermo Fisher Scientific), setting the following thresholds and methods: Gene-Level Fold Change < -2 or > 2, Gene-Level P-Value < 0,05, Splicing Index < -2 or > 2, Exon-Level P-Value < 0,05, Anova Method: ebayes, Probeset (Gene/Exon) considered expressed if ≥ 50% samples have DABG values below DABG Threshold (DABG < 0,05), event Pointer P-Value < 0,1 and event score > 0,2. Data are deposited in GEO repository with number GSE108007.

RNA-seq experiments and analysis

An aliquot of the same RNA extracted from COS-7 cells for splicing arrays (previous section) was used for RNASEQ, in three biological replicates each: control, (CGG)60X 185ng, siTRA2A 50nM, (CGG)60X 185ng+siTRA2A 13.6ng, GFP-TRA2A 200ng, (CGG)60X 185ng + GFP-TRA2A 200ng. 500ng of total RNA were used for library preparation with TruSeq total RNA-rRNA depletion (Illumina). The sequencing was performed with HiSeq3000, paired-end, 2X125, 3samples/lane.

Nucleotide alignments were performed using STAR_2.5.2 with reference genome and annotations taken from Gencode Release 27 (GRCh38.p10) (http://www.gencodegenes.org/releases/current.html. Pre-processing of bam files was made with python scripts ("dexseq_prepare_annotation.py" and "dexseq_count.py") provided in the DEXSeq v 1.24.2 R package. Statistical analysis of alternative splicing events was done with R 3.4.0 using two methods: EventPointer v1.0.0 and DEXSeq v 1.24.2.

Supplementary Figures and Tables.

Supplementary Figure 1. Datasets A) Granule RBPs Red circle: granule-forming proteins, Blue circle: RBPs, as defined in Gerstberger et al, 2014 (7). Intersection represents granule RBPs. **B)** Number of interactions. Red circle: granule-forming proteins. Blue circle: RBPs with known targets. Intersection represents granule RBPs with known targets.

Supplementary Figure 2. Distribution of centrality values of granule and non-granule RBPs in different interaction networks. A) Centrality distributions for the human dataset. Up: Protein-protein network. (p-value (left) = 0.39, p-value (centre) = 0.41, p-value (right) = 0.36. Down: Protein-RNA network (p-value (left) = 0.003, p-value (centre) = 0.007, p-value (right) = 0.01. **B)** Centrality distributions for the yeast dataset. Up: Protein-protein network. (p-value (left) = 0.26, p-value (centre) = 0.30, p-value (right) = 0.18. Down: Protein-RNA network (p-value (left) = 0.02, p-value (centre) = 0.05, p-value (right) = 0.01.

Supplementary Figure 3. Number of RNA targets of granule and non-granule RBPs: A) First quartile of the reads/expression distribution (Q1). **B)** Second quartile (Q2).

Supplementary Figure 4. Properties of granule RNAs. A) RNAs interacting exclusively with granule forming RBPs have higher number of protein contacts (p-value = 0.04, Wilcoxon test). Human transcripts: **B)** Granule RNAs have more structured UTRs (p-value = 0.007; KS test). PARS analysis on 3'UTR of granule and non-granule RNAs**.** Yeast granule RNA are **C)** structured (p-value = 0.001; KS test; PARS data), and **D)** more abundant (p-value = 2.2e-16; KS test) than non-granule RNAs. The UTR analysis was not performed due to the lack of annotation.

Supplementary Figure 5. Computational predictions of granule-forming components. A) Granule transcripts are predicted to be more structured. Structural content according to CROSS is higher in granule RNAs (p-value < 2.2e-16, KS test). B and C) *cat*GRANULE performances on human and yeast experimentally described granule-forming proteins. AUC (Area under the ROC curve) is used to measure the discriminative power of the method. D) Distribution of *cat*GRANULE scores for the whole human proteome. TRA2A (*cat*GRANULE score = 2.14) ranks 188th out of 20190 human proteins (i.e. 1% of the distribution).

Supplementary Figure 6. TRA2A levels in human lymphocytes and COS-7 cell model. A) Human lymphocytes from control (A) or pre mutation-carrier (B) were lysated and both RNA and protein were isolated (*** p-value < 0.01). Relative TRA2A RNA expression (left panel) and TRA2A protein (right panel) are represented. **B)** COS-7 cells were transfected with CGG(60X) and compared to controls. After 24h, 48h or 72h of transfection cells were pelleted and RNA and protein extraction was performed. Relative TRA2A RNA expression (left panel) and TRA2A protein (right panel) are represented.

Supplementary Figure 7. TRA2A over-expression and TRA2B knock-down. A) Control COS-7 cells were transfected with siTRA2B and GFP-TRA2A (in absence of CGG(60X) transfection). **B)** COS-7 cells were transfected with CGG(60X), siTRA2B and GFP-TRA2A. In both A and B, after 48 hours of transfection cells were hybridized with Cy3-GGC(8X) probe and immunostained with antiGFP. The graphs represent TRA2A/CGG levels.

Supplementary Figure 8. TRA2A and TRA2B over-expression COS-7 cells were transfected with GFP-TRA2A **A)** or GFP-TRA2B **B)** and CGG(60X). After 48 hours, cells were hybridized with Cy3-GGC(8X) probe and immunostained with an antibody against either TRA2A or TRA2B. Graphs represent TRA2A/TRA2B/CGG levels.

Supplementary Figure 9. TRA2B over-expression and TRA2A knock-down. A) Control COS-7 cells (without CGG(60X) transfection) were transfected with GFP-TRA2B and siTRA2A. **B)** COS-7 cells were transfected with CGG(60X), GFP-TRA2B and siTRA2A. In both A and B, after 48 hours, cells were hybridized with Cy3-GGC(8X) probe and immunostained with an antibody against TRA2B. The graph represents TRA2B/CGG levels.

Supplementary Table 1. Protein and RNA interactions. A) list of human and yeast granule proteins. **B, C, D**) RNA partners of human RBPs identified at different cut-offs of the reads/expression distribution (first, second and third quartile are indicated with Q1, Q2 and Q3). Names starting by NM indicate coding transcripts and names starting by NR indicate non-coding transcripts. **E)** RNA partners of yeast RBPs. **F, G, H)** Overlap between interactomes of human RBPs calculated using the Jaccard index (first, second and third quartile are indicated with Q1, Q2 and Q3) and **I)** Overlap between yeast RBP interactomes.

Supplementary Table 2. A) RBP contacts of human RNAs. Names starting by NM indicate coding transcripts and names starting by NR indicate non-coding transcripts. **B)** Number of total, granule and non-granule contacts, structural content, length and UTR size of human transcripts; **C)** Number of total, granule and non-granule contacts, structural content, length of yeast transcripts.

Supplementary Table 3. *in silico* predictions of CGG interactions. *cat*RAPID scores (discriminative power *DP*, interaction strength *IS*) (11), name of the gene, catGRANULE score (16), granule ability (predicted / validated) and empirical p-value indicating the ability of proteins to interact with CGG repeats (calculated on 3340 DNA-binding, RNA-binding and structurally disordered proteins).

Supplementary Table 4. *in vitro* validation of CGG interactions. We employed protein arrays to perform a large *in vitro* screening of RBP interactions with the first *FMR1* exon (17, 18). We probed both expanded ("PRE"; 79 CGG) and normal ("WT"; 21 CGG) on three independent arrays, obtaining highly reproducible results.

Supplementary Table 5. Microarray and RNA-seq analysis of splicing events. A) CGG over-expression vs CTL; B) CGG over-expression and TRA2A knock-down vs control. Fold Changes, significance and (sub)-exon names are reported. Microarray: Transcriptome Analysis Console (TAC) 4.0 software (ThermoFisher) was used to identify splicing events. RNA-seq: Statistical analysis of alternative splicing events was done using EventPointer v1.0.0 and DEXSeq v 1.24.2 (GO analysis: http://www.tartaglialab.com/GO_analyser/render_GO_universal/2105/64ce4f8d1d/, http://www.tartaglialab.com/GO_analyser/render_GO_universal/2108/eef220536a/).

1. Jain S, et al. (2016) ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* 164(3):487–498.

2. Huttlin EL, et al. (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162(2):425–440.

3. Van Nostrand EL, et al. (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 13(6):508–514.

4. Armaos A, Cirillo D, Gaetano Tartaglia G (2017) omiXcore: a web server for prediction of protein interactions with large RNA. *Bioinformatics*. doi:10.1093/bioinformatics/btx361.

5. Mittal N, Scherrer T, Gerber AP, Janga SC (2011) Interplay between Posttranscriptional and Posttranslational Interactions of RNA-Binding Proteins. *J Mol Biol* 409(3):466–479.

6. Brannan KW, et al. (2016) SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol Cell* 64(2):282–293.

7. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. *Nat Rev Genet* 15(12):829–845.

8. Nepusz T CG The igraph software package for complex network research. *InterJournal Complex Syst* 1695. 2006.

9. Wan Y, et al. (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505(7485):706–709.

10. Kertesz M, et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467(7311):103–107.

11. Agostini F, et al. (2013) catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinforma Oxf Engl* 29(22):2928–2930.

12. Bellucci M, Agostini F, Masin M, Tartaglia GG (2011) Predicting protein associations with long noncoding RNAs. *Nat Methods* 8(6):444–445.

13. Bolognesi, Benedetta, et al. Dosage sensitivity caused by increased protein concentration triggering a liquid phase. *Cell Rep* 16:10.1016/j.celrep.2016.05.076.

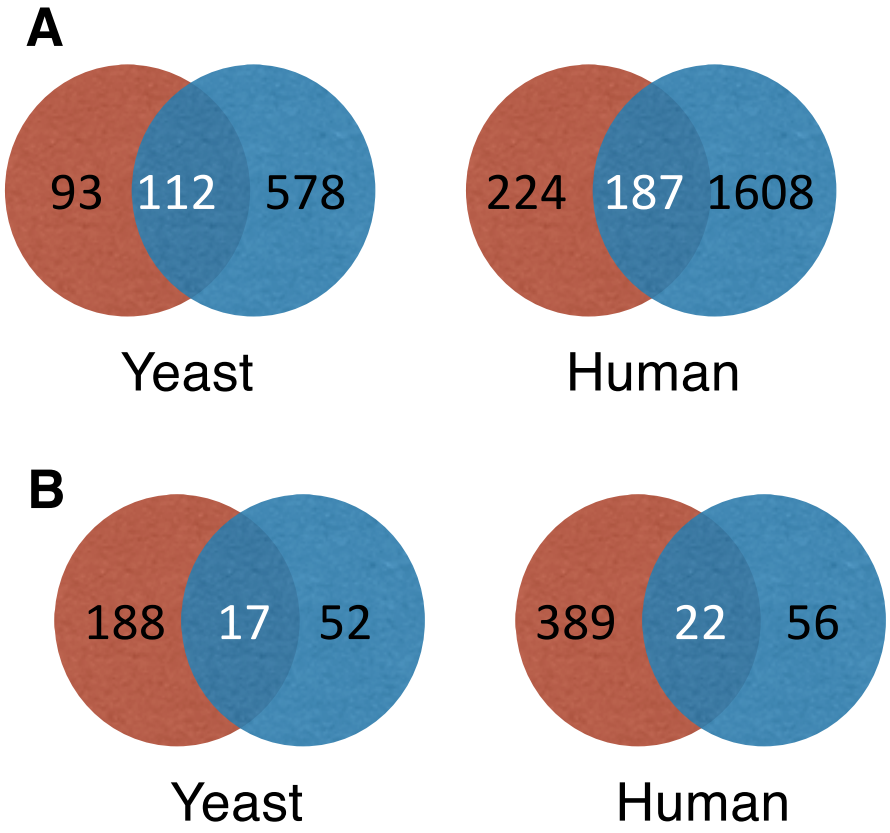
14. Mateu-Huertas E, et al. (2014) Blood expression profiles of fragile X premutation carriers identify candidate genes involved in neurodegenerative and infertility phenotypes. *Neurobiol Dis* 65:43–54.

15. Buijsen RA, et al. (2014) FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun* 2:162.

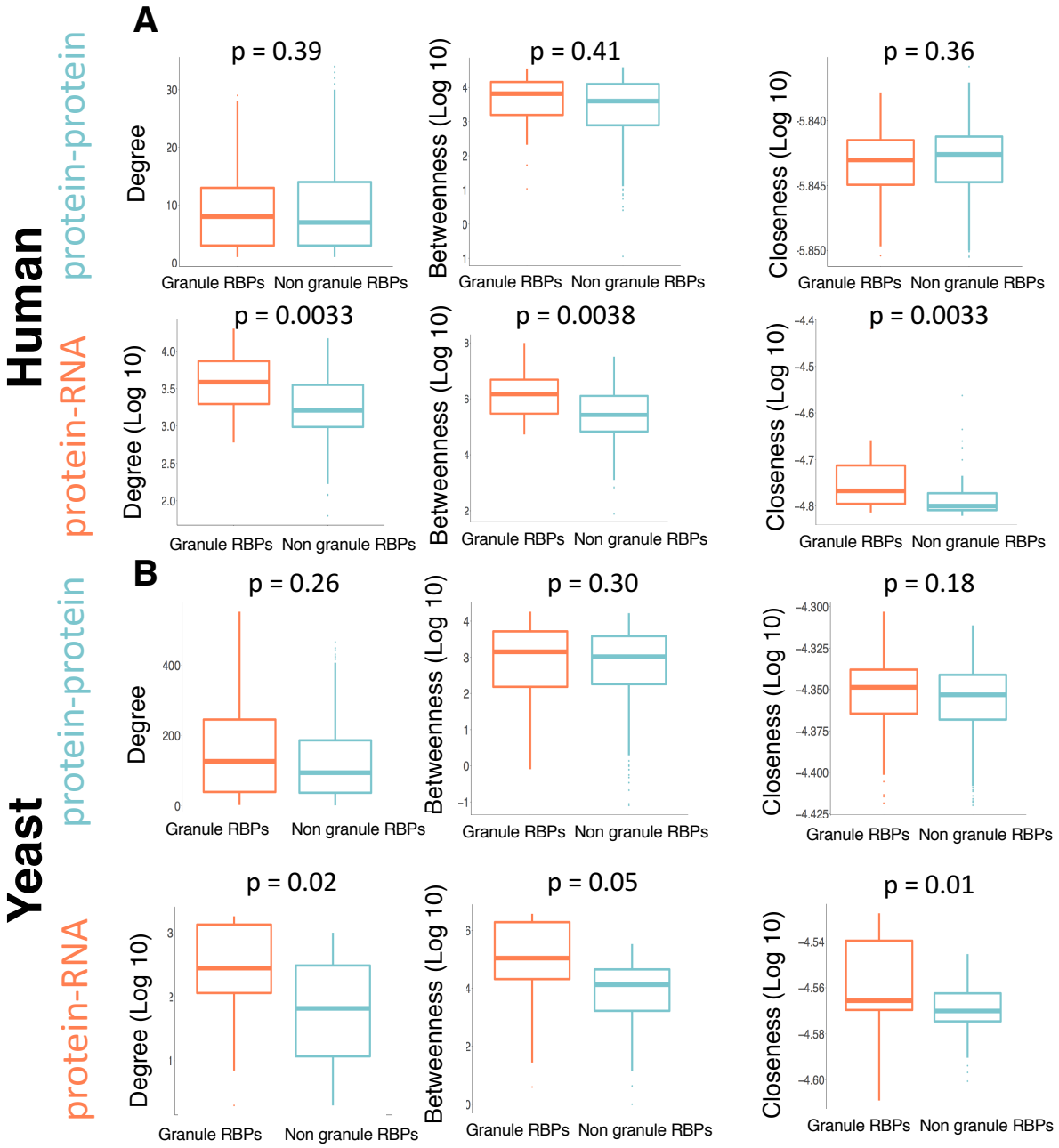
16. Bolognesi B, et al. (2016) A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep* 16(1):222–231.

17. Marchese D, et al. Discovering the 3′ UTR-mediated regulation of alpha-synuclein. *Nucleic Acids Res*. doi:10.1093/nar/gkx1048.

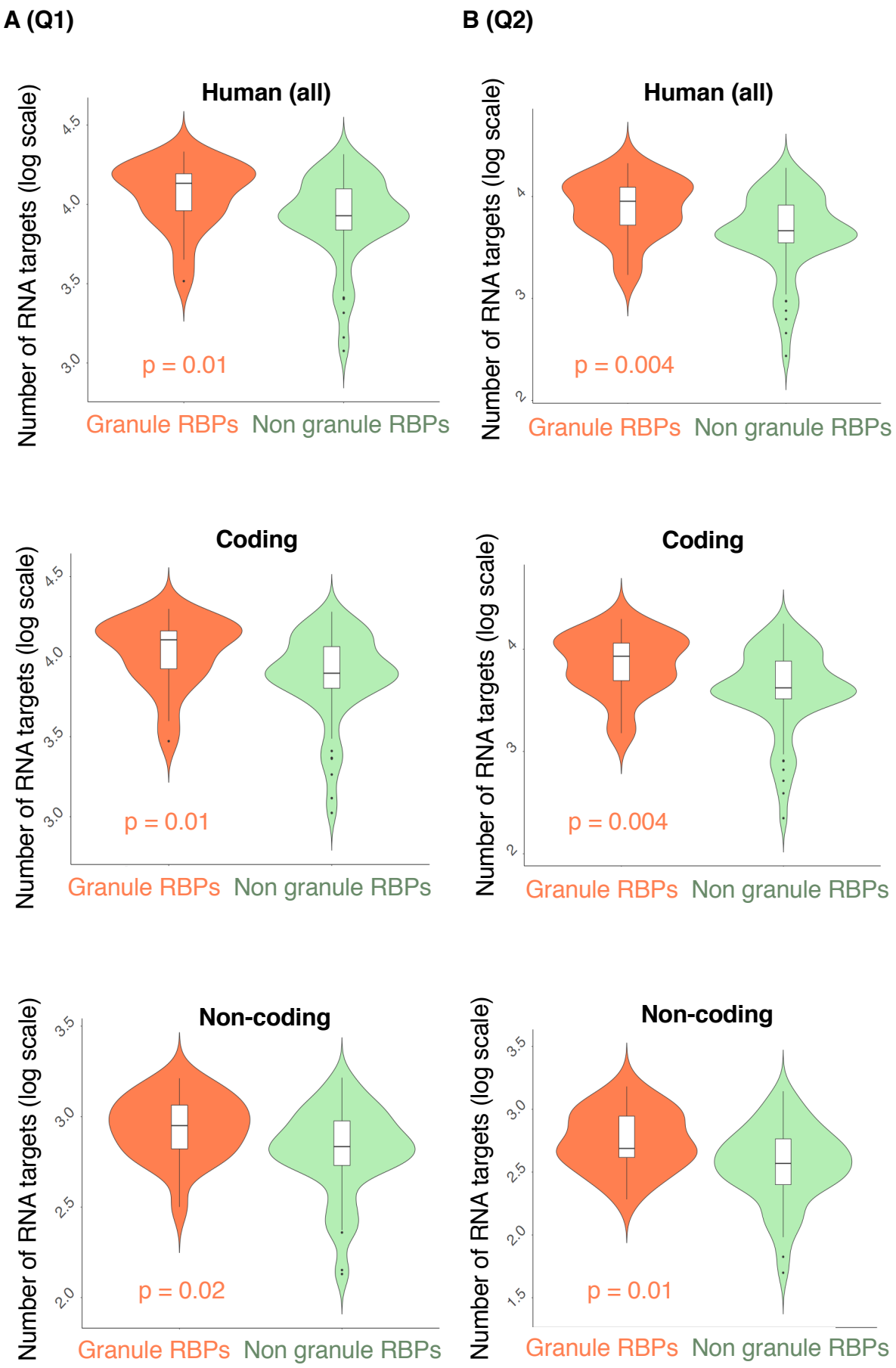
18. Cirillo D, et al. (2017) Quantitative predictions of protein interactions with long noncoding RNAs. *Nat Methods* 14(1):5–6.



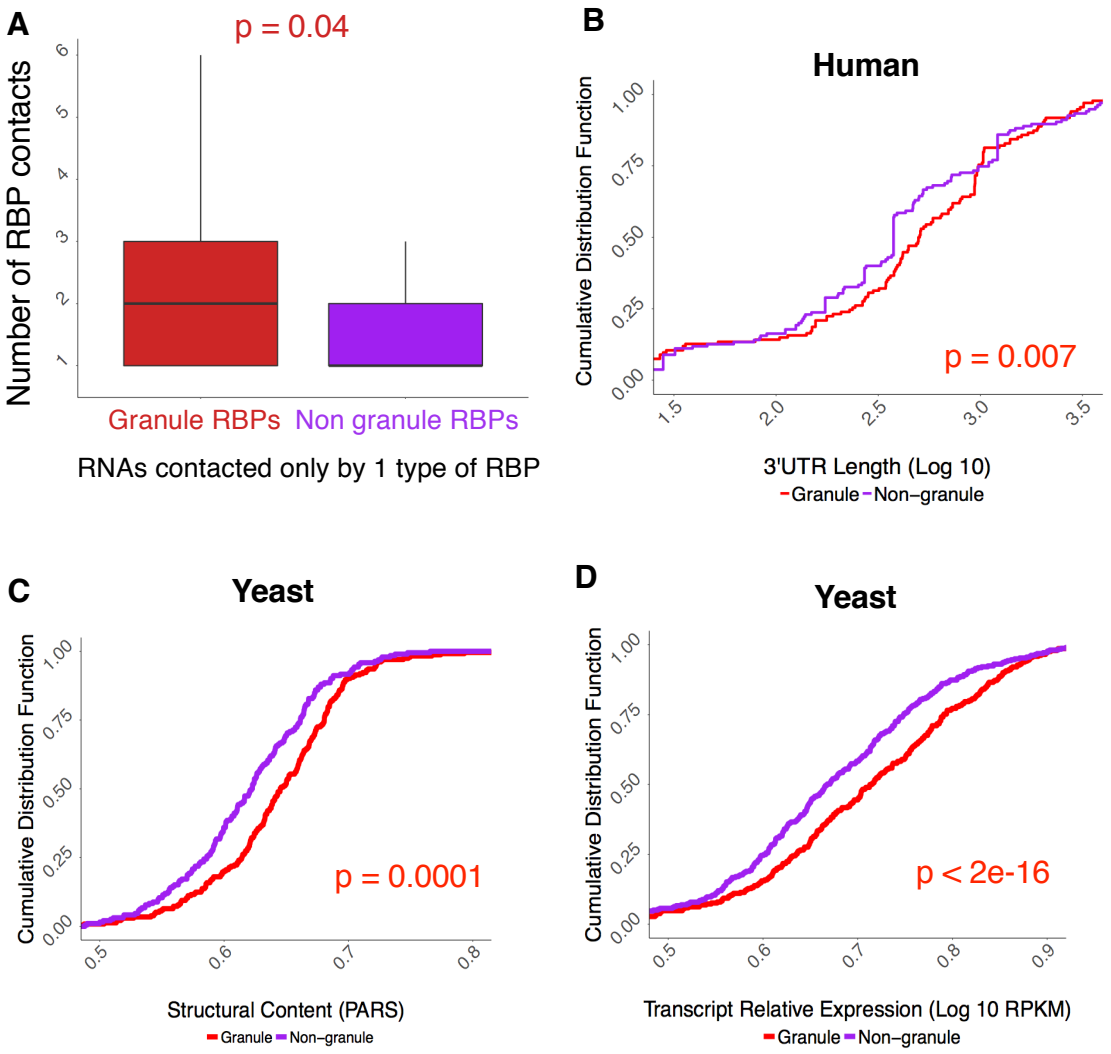
Supplementary Figure 1



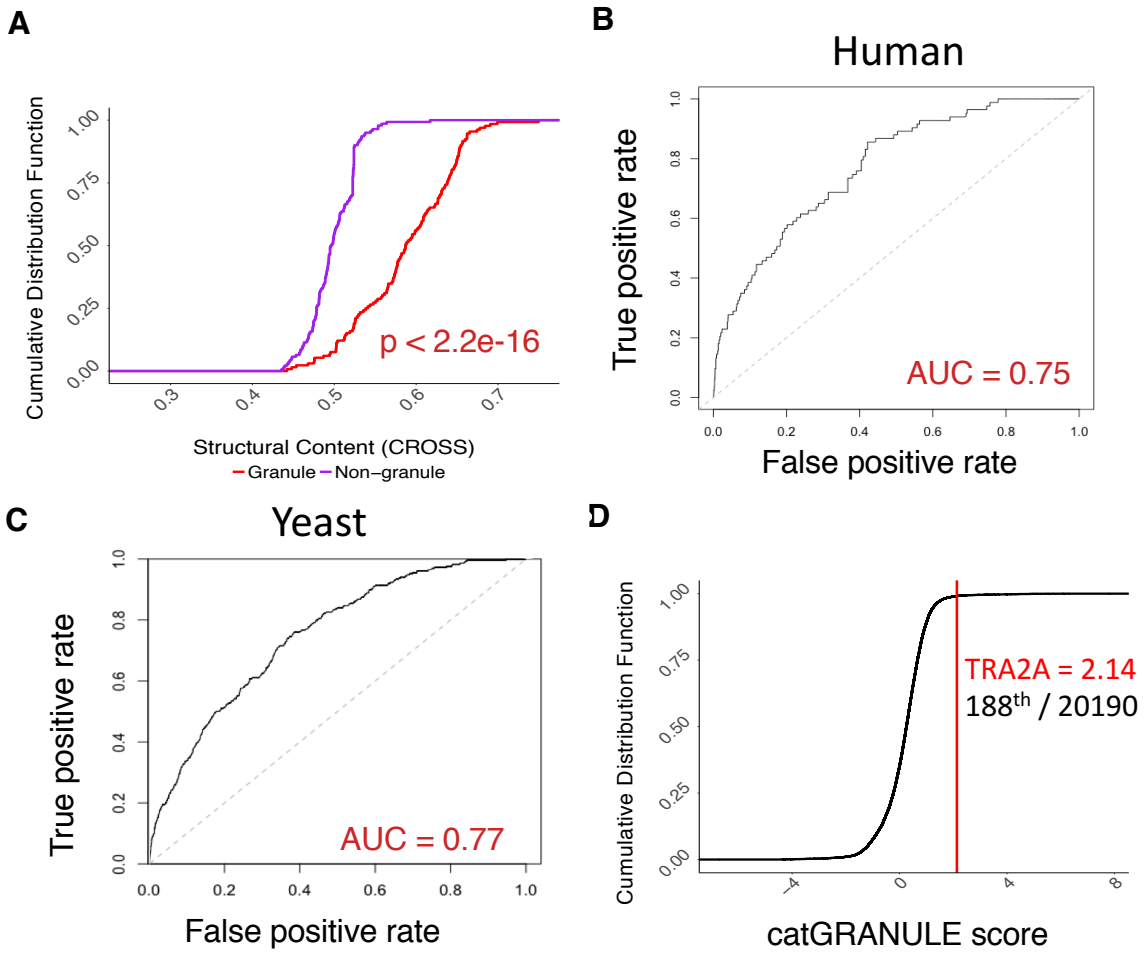
Supplementary Figure 2



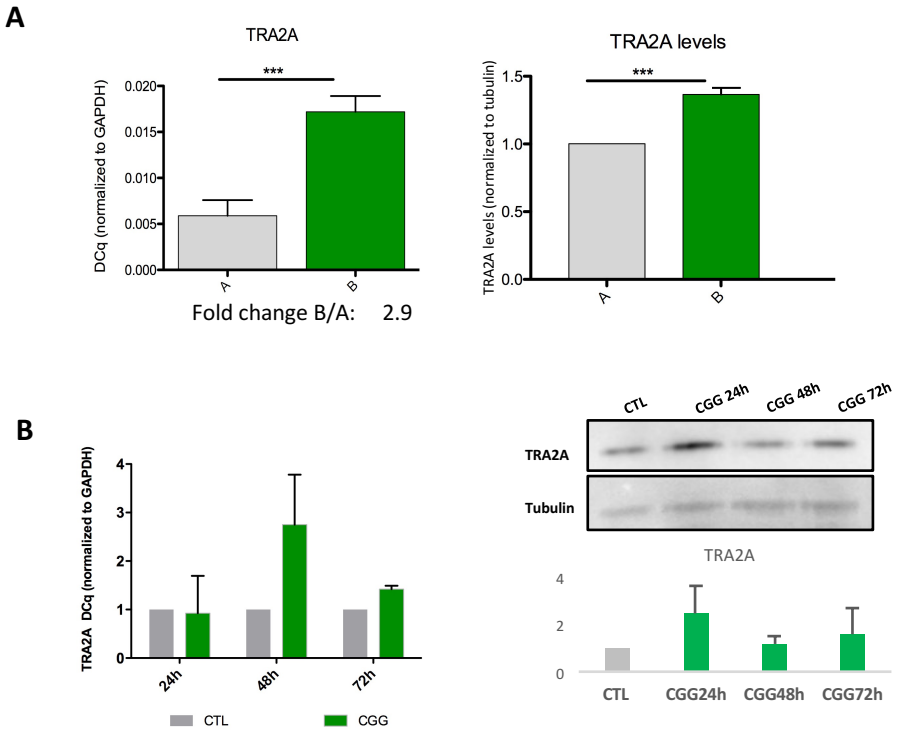
Supplementary Figure 3



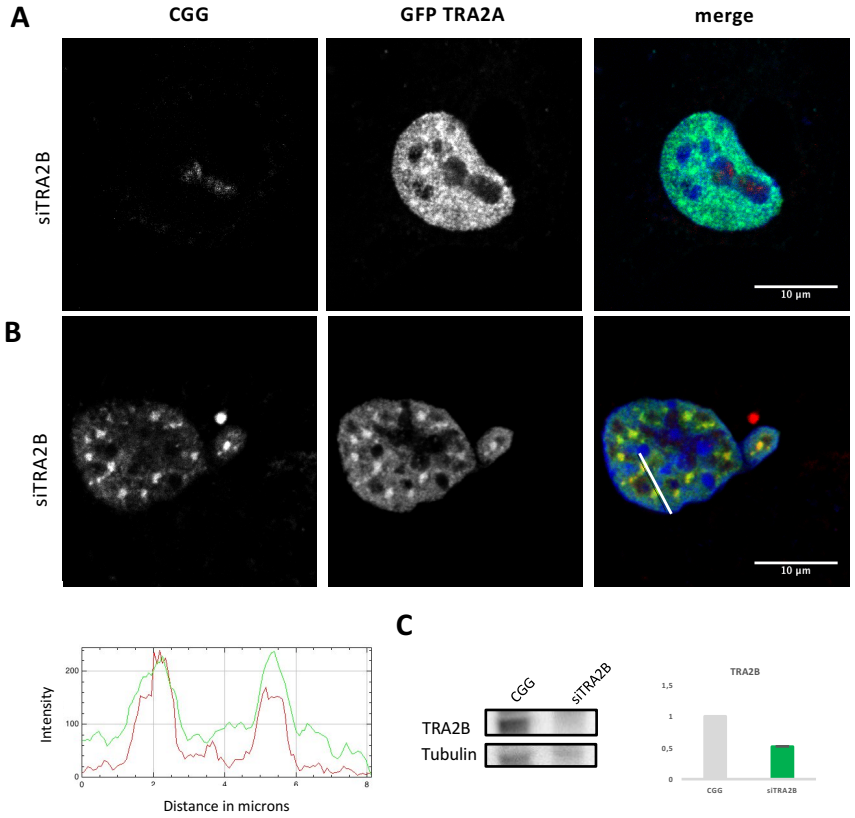
Supplementary Figure 4



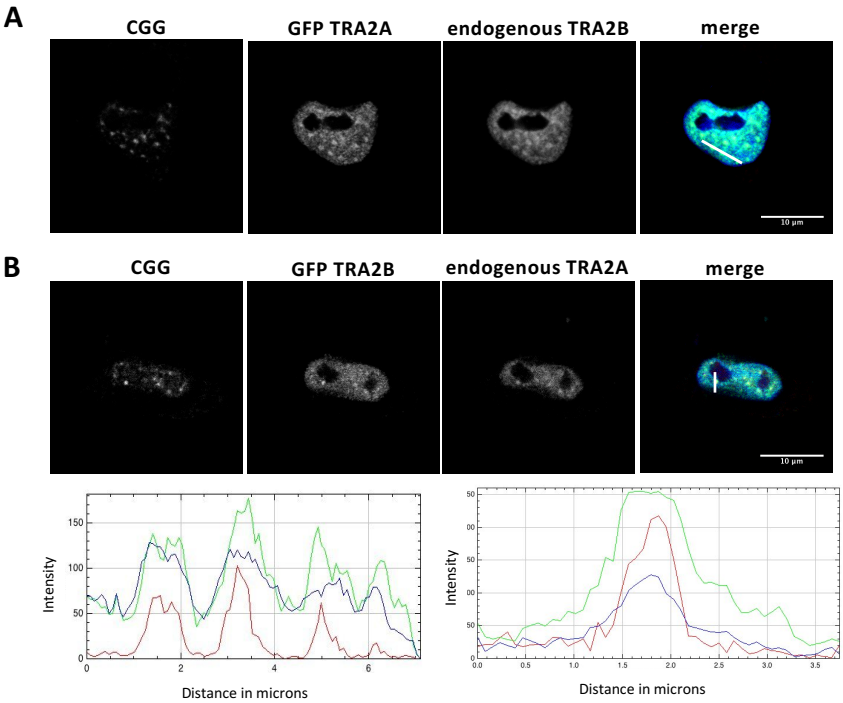
Supplementary Figure 5



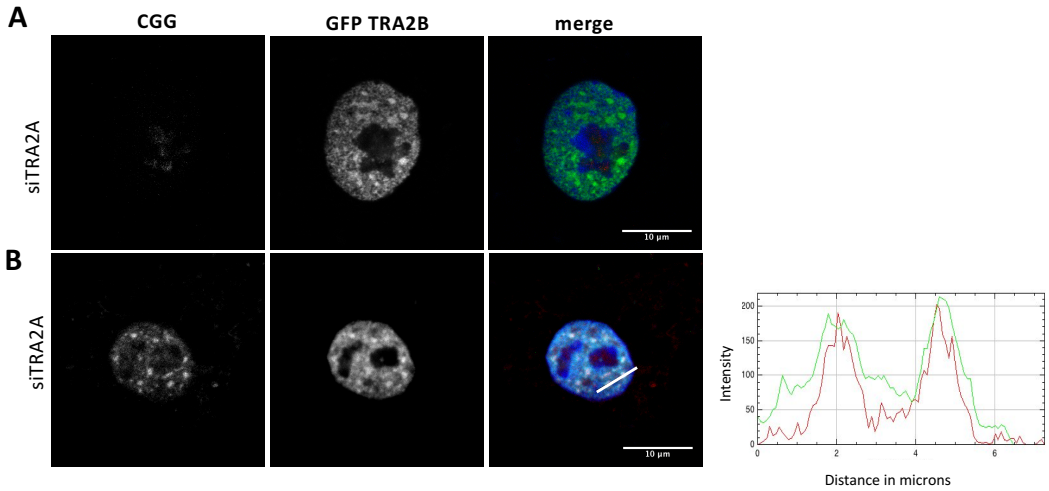
Supplementary Figure 6



Supplementary Figure 7



Supplementary Figure 8



Supplementary Figure 9