Phase Separations Driven by RNA Scaffolds and Protein Sequestration in FXTAS

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 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 aggregating 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 phase-separations in the cell by acting as scaffolds for protein assembly.

In a proof-of-concept experiment, 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 the sequestration of a previously unreported partner, TRA2A, and consequent impairment of splicing pathways associated with mental retardation and intellectual disability.

Introduction

Ribonucleoprotein (RNP) aggregates or granules are liquid-like cellular compartments composed of RNA-binding proteins (RBPs) and transcripts[1] that assemble through a process of phase separation and are in dynamic exchange with the surrounding environment[2]. RNP granules such as processing bodies and stress granules (SGs) are evolutionarily conserved from yeast to human[3] and contain constitutive protein components such as G3BP1 (yeast: Nxt3), TIA1 (Pub1), and TIAR (Ngr1)[4]. Several RBPs involved in physiological liquid-liquid phase separations are prone to form amyloid aggregates if subjected to amino acid mutations[5,6] that indice a transition from a liquid droplet to a solid phase[7]. 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[8] and Myotonic Dystrophy[9].

Recent evidence indicates that RBPs act as scaffolding elements promoting RNP granule assembly through protein-protein interactions (PPI)[10,11], 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[12]. If distinct RNA species are present in SGs, a fraction of them could be involved in mediating RBP sequestration. In this regards we recently observed that different RNAs can act as scaffolds for RNP complexes[13], which leads to the intriguing hypothesis that the transcripts could be also 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)[14] and mass spectrometric analysis of SGs[3], we identified a class of transcripts that bind to a large number of proteins and, therefore, qualify as potential scaffolding elements. Our computational analysis reveals that untranslated regions (UTRs) have a particularly strong potential to attract protein complexes in RNP granules, especially when containing specific homo-nucleotide repeats. In agreement with this finding, 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[15,16].

In a proof-of-concept experiment, we focused on 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[17]. The premutation

range of *FMR1* (55–200 CGG repeats) is accompanied by appearance of intranuclear aggregates (*FMR1* inclusions) that are the typical hallmark of Fragile X-associated Tremor / Ataxia Syndrome (FXTAS)[17].

Two mechanisms have been suggested to explain FXTAS onset and development[18]: 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)[17]. Previous work indicates that *FMR1* inclusions contain specific proteins such as HNRNP A2/B1, MBNL1, LMNA and INA[19]. Other proteins such as CUGBP1, KHDRBS1, DGCR8, which are involved in splicing regulation, mRNA transport and regulation of microRNAs[20–23] as well as FMRpolyG peptides have been found in the inclusions[24].

The lability of *FMR1* inclusions, which makes them not suitable for biochemical purification[25,26], does not allow to easily distinguish between direct and indirect partners of *FMR1* 5' UTR and there is need of novel strategies to characterize their protein composition. Moreover, as current clinical strategies are limited to palliative care to ameliorate specific FXTAS symptoms there is insufficient knowledge of targets for therapeutic intervention[17,24] and a strong demand for innovative approaches to identify molecules recruited in *FMR1* inclusions[27].

We studied the scaffolding ability of *FMR1* 5' UTR using a novel *in silico* / *in vitro* approach to reveal physical interactions[28,29]. In addition to retrieving known *FMR1* 5' UTR partners, we discovered previously unreported RBPs, including the splicing regulator TRA2A. Analysis of TRA2A recruitment in *FMR1* inclusions revealed splicing patterns affected in genes associated with mental retardation and intellectual disabilities. We corroborated our results by showing TRA2A aggregation in a FXTAS mouse model as well as *post mortem* brain tissues from FXTAS patients.

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 (**Online Material** and **Supplementary Table 1A**). The protein-RNA interaction datasets were identified through eCLIP (human)[14] and microarray (yeast)[30] 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[30,31] **(Online Material)**[30,31]. 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 centrality measures of the granule network **(Figure 1A** and **Supplementary Figure 2)**. Importantly, both yeast and human granule RBPs interact with a larger number of 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; **Online Material**). Such a difference holds 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[32]), 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**)[12].

Non-coding RNAs are present in RNP granules

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 described as an architectural RNA implicated in paraspeckle nucleation[33]. We hypothesize that other highly contacted long non-coding RNAs may have similar functions within cytoplasmic RNP granules. NORAD, a recently described long non-coding RNA involved in genomic stability, also interacts with all the proteins in our dataset[34]. NORAD has repetitive sequence units, activation upon stress and ability to sequester proteins[35], 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 (**Online Material**), we found that RNAs enriched in granule RBP contacts are more expressed (**Figure 2A**; p-value = 5e-11, KS test; **Online Material**), 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 work reporting how length[36], structure[37] and abundance[38] contribute to RNA assembly into RNP granules[12].

The increase in structural content is significant in the 5' UTRs of granule-associated transcripts (**Figure 2D**; p-value = 0.04; KS test) and their composition is enriched in trinucleotides with ability to self-assemble into hairpin structures[39]: occurrence of

CCG, UGC, CGC, GGU and CGG triplets significantly discriminate granule and non-granule transcripts with (AUCs > 0.60; **Figure 2E**). In agreement with these findings, CROSS[40] predictions of RNA structure indicate that double-stranded regions are enriched in granule-associated transcripts (**Supplementary Figure 5A** and increase proportionally to the length of CGG repeats (**Figure 2F**), which is also in line with UV-monitored structure melting experiments[39].

In silico predictions indicate a large number of CGG partners

To further investigate the scaffolding ability of homo-nucleotide expansions, we selected the *FMR1* transcript containing CGG repetitions. Using *cat*RAPID *omics* (**Online Material**)[41], we computed interactions between the 5' *FMR1* UTR (containing 79 CGG repeats in the premutation range) and a library of nucleic-acid binding proteins (3340 DNA-binding, RNA-binding and structurally disordered proteins[42]). Previously identified CGG-binding proteins[21] such as ROA 1, 2 and 3, HNRNP C, D and M, SRSF 1, 4, 5, 6, 7 and 10 as well as MBNL1 and KHDRBS3 are 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 found for a set of 92 RBPs that have been reported to assemble in SGs[3] (**Supplementary Table 3**). Our calculations identify an extra group of 37 RBP interactions that are predicted to form granules by the *cat*GRANULE algorithm[2] (**Online Material; Supplementary Figures 5B and 5C**). Among the RBPs prone to physically associate with *FMR1*, we predicted 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

In parallel, we employed protein arrays to perform a large *in vitro* screening of RBP interactions with the first *FMR1* exon[28,29]. We probed both expanded (79 CGG) and normal (21 CGG) range repeats on three independent arrays, 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)[29,43].

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 (5%) and MBNL1 (10%). 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 ‰) we identified 27 previously unreported interactions (binding to expanded and normal range repeats). The most significant GO

term in the list of 85 RBPs relates to splicing activity (FDR $<10^{-7}$, as reported by GeneMANIA[44] and *clever*GO[45]) and includes SRSF 1, 5, 6 and 10, PCBP 1 and 2, HNRNP A0 and F, NOVA1, PPIG and TRA2A. Notably, *cat*RAPID *omics* predictions show strong agreement with protein array experiments: from low- to high-affinity interactions, *cat*RAPID performances increase progressively reaching AUCs of 0.80 (**Figure 3C**). Two previously predicted interactors[43], CIRBP and PTBP2, rank in the top 2% of all fluorescence intensities.

Out of 27 candidates (signal to background ratio > 2.5), 24 were predicted by *cat*GRANULE[2] to form granules and among them the splicing regulator TRA2A showed the highest score (granule propensity = 2.15; **Figure 3D**; **Supplementary Figure 5D; Supplementary Table 3**). Their specificity for CGG repeats was also checked in relation with binding to SNCA 3' UTR (**Figure 3D**)[28,29]. In agreement with our predictions, eCLIP experiments indicate that the *FMR1* transcript ranks in the top 25% of strongest interactions of TRA2A[14].

As splicing defects have been reported to occur in FXTAS disease[18,21], we decided to prioritize TRA2A for further experimental characterization.

TRA2A recruitment by CGG inclusion is confirmed by an in vivo model

We 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 (primate) to study *FMR1* inclusions[21]. Transfection of a plasmid containing CGG repeats (60X) induce significant increase in both RNA and protein levels of TRA2A after 48 hours (**Supplementary Figure 6**)[21]. By means of RNA FISH coupled to immunofluorescence (**Online Material**), we found that CGG expansions form nuclear inclusions that co-localize with endogenous TRA2A (**Online Material**)[21]. TRA2A shows instead a diffuse nuclear pattern in cells that do not over-express CGG repeats (**Figure 4A**).

Upon knockdown of TRA2A using siRNA (**Online Material**) 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. Using TMP4yP to specifically unfold CGG repeats[46] , we found that the aggregates are disrupted and TRA2A remains diffuse

(**Figure 5A**). 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 RBP interactions, we blocked TRA2A sequestration (**Figure 5B**). Our latter experiment indicates that the stability of RNA structure is essential for TRA2A recruitment, as suggested by our initial hypothesis (***Characteristic features of candidate scaffolding RNAs;*** **Figure 2F**).

TRA2A recruitment in RNA inclusions is independent of TRA2B

Using RNA FISH coupled to immunofluorescence, we found that a TRA2A partner, TRA2B (not present in the protein array)[44], forms inclusions when COS-7 cells are transfected with CGG repeats, in agreement with previous *in vitro* screenings[12], (**Figure 6A**). The same result is observed upon TRA2A knockdown (**Figure 6B)**.

Upon TRA2B knockdown, endogenous TRA2A is recruited by CGG inclusions (**Figure 6B;** the same result is observed with over-expressed TRA2A; **Supplementary Figure 7**). In absence of CGG and TRA2B instead, over-expressed TRA2A do not localize into inclusions (**Supplementary Figure 7A**). Upon TRA2B knockdown but in presence of CGG, over-expressed TRA2A is still 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, we conclude that TRA2A and TRA2B are independently recruited by CGG RNA aggregates.

Effects of TRA2A recruitment on RNA splicing

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

Splicing alterations due to TRA2A sequestration were studied through two independent techniques, microarrays and RNA-seq (both in triplicate experiments; **Online Material**), which we used to investigate events (i) occurring upon CGG repeats over-expression and (ii) affected when TRA2A is knocked-down. Applying these criteria, we identified 53 exons subjected to alternative splicing (36 skipped and 17 included exons; q-value<0.1; **Figure 7A; Supplementary Table 5**). Using the *clever*GO algorithm[45] we found that the largest GO cluster of affected genes

includesRBPs (18 genes; **Supplementary Table 5**; 'RNA-binding'; fold enrichment of 24; p-value < 10^{-8}; calculated with Bonferroni correction; examples: HNRNPL, CIRBP and DDX24), spliceosome components ('mRNA splicing via spliceosome'; fold enrichment of 5; p-value < 10^{-3}, examples: ROA 2 and SRSF 10) and other 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[48], ACTB[49] and ACTG1[50], have splicing patterns affected by TRA2A sequestration. Similarly, muscle related proteins, such as PIP5K1C[51], TPM1[52] and genes linked to intellectual disabilities, including DOCK3[53], are subjected to exon skipping upon TRA2A aggregation (**Supplementary Table 5; Figure 7B**).

Thus, 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[17]. The main RAN translation product, FMRpolyG, co-localizes with ubiquitin in intranuclear inclusions, as observed in FXTAS human brain and mouse models[17,24,54].

We tested if TRA2A co-aggregates with FXTAS inclusions and FMRpolyG in a mouse model in which the *FMR1* 5'UTR (90 CGG repeats) was expressed under the control of doxycycline leading to formation of ubiquituin and FMRpolyG positive intranuclear inclusions[55]. 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**), co-localizing with FMRpolyG (**Figure 8C**).

Thus, TRA2A sequestration by CGG repeats is not only observed in cell lines, but also in FXTAS animal models and post-mortem 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[56]. Yet, based on our analysis of interaction networks, we propose that specific RNAs can serve as structural scaffolds to recruit proteins in RNP granules. We predict that scaffolding RNAs are characterized by a large number of RBP contacts, increased length and structural content.

To experimentally validate our findings we focused on the CGG repeat expansions present in the *FMR1* 5' UTR that assemble in hairpin structures triggering formation of RNP inclusions observed in FXTAS[18,21]. Combining *in silico* predictions of protein-RNA interactions with large-scale *in vitro* validations[28,29], 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.

Using human primary tissues as well as primate and mouse models of FXTAS, we further characterized TRA2A sequestration in *FMR1* inclusions. TRA2A is a splicing regulator containing Ser/Arg repeats near its N- and C-*termini*[57]. Similarly to other splicing factors, TRA2A contains structurally disordered regions and multiple stretches of low sequence complexity[58] that promote liquid-liquid phase separation[2]. Although formation of large assemblies is crucial for splicing activity and regulation[59,60], liquid-liquid phase separations have been mainly observed in the context of disease. Indeed, aggregation of DMPF and ZNF9 transcripts (containing CUG or CCUG expansions) has been reported to induce sequestration of MBNL1 and partners[9] leading to impairment of splicing activity and Myotonic Dystrophy[61].

In agreement with our predictions, we found that TRA2A is sequestered by nuclear RNA inclusions upon over-expression of *FMR1* 5' UTR in a cellular model of FXTAS. As identified in our *in silico* analysis, PPI networks do not play a direct role in protein recruitment: both endogenous and overexpressed TRA2A are recruited to CGG granules upon knockdown of its partner TRA2B[47], and the assembly process is only driven by RNA inclusions formation. Unfolding the structure of CGG repeats through TMP4yP[46] or using 9-hydroxy-5,11-dimethyl-2-(2-(piperidin-1-yl)ethyl)-6H-pyrido[4,3-b]carbazol-2-ium[20] to block protein binding 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, such as ACTB[49] and ACTG1[50], intellectual disabilities, including

DOCK3[53] and muscle-related protein WWP2[62,63], which is relevant in the context of FXTAS.

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 these 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*[17] was not present in our system, when using brain samples from FXTAS post-mortem donors we observed that TRA2A co-localizes with the polypeptide FMRpolyG in brain inclusions. Thus, our results indicate that the 5' UTR of *FMR1* acts as a scaffold also for FMRpolyG.

In conclusion, we identified the characteristics that make RNAs scaffolding elements for the formation of RNP granules. Our computational analysis highlighted specific features: the number of protein contacts, increased secondary structure content, abundance and UTR length. In a proof-of-concept, we showed that the *FMR1* 5' UTR acquires the ability to scaffold RNP granules upon a disease-related CGG repeat expansion, which increases UTR length and structural content. The *in silico*, *in vitro* and *in vivo* investigations lead to the identification of RBPs involved in alternative splicing, such as TRA2A, which we showed to be specifically sequestered into RNP inclusions of FXTAS. Since sequestration of TRA2A in *FMR1* inclusions causes alterations in downstream splicing events, we envisage that TRA2A and related targets could be used as biomarkers for prognostic and diagnostic purposes.

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[3] (Area under the ROC curve AUC is used to measure the enrichment in SGs; **Online Material**). 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-val < 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[40] is employed to measure the amount of double-stranded RNA).

Figure 3. Protein interactions of CGG repeats. A) Using *cat*RAPID *omics*[41], 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)[21]. 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[28]. 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*[41] 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)[2] and the splicing regulator TRA2A (red bar) shows the highest propensity. The black bars indicate non-specific partners (interacting with SNCA 3' UTR[28,29] or showing poor RNA-binding propensity[42]).

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. A) COS-7 cells were co-transfected with empty vector or CGG(60X), and after 24h of transfection cells were treated with TMP4yP. **B)** COS-7 cells were treated similarly as in (A) but with 1a molecule instead of TMP4yP. 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; Online Material**). **Inset**: GO

analysis of the largest cluster (18 genes) affected by TRA2A sequestration[45] reveals links to RBP activities and splicing activities. **B**) Exon inclusion/skipping of genes associated with mental retardation, such ACTB[49] and ACTG1[50], intellectual disabilities, including DOCK3[53] and muscle-related protein WWP2[62,63] 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 (**Online Material**).

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[55,65]. 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*) [41], name of the gene, catGRANULE score[2], 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 [22,37]).

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 [28,29]. 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/).

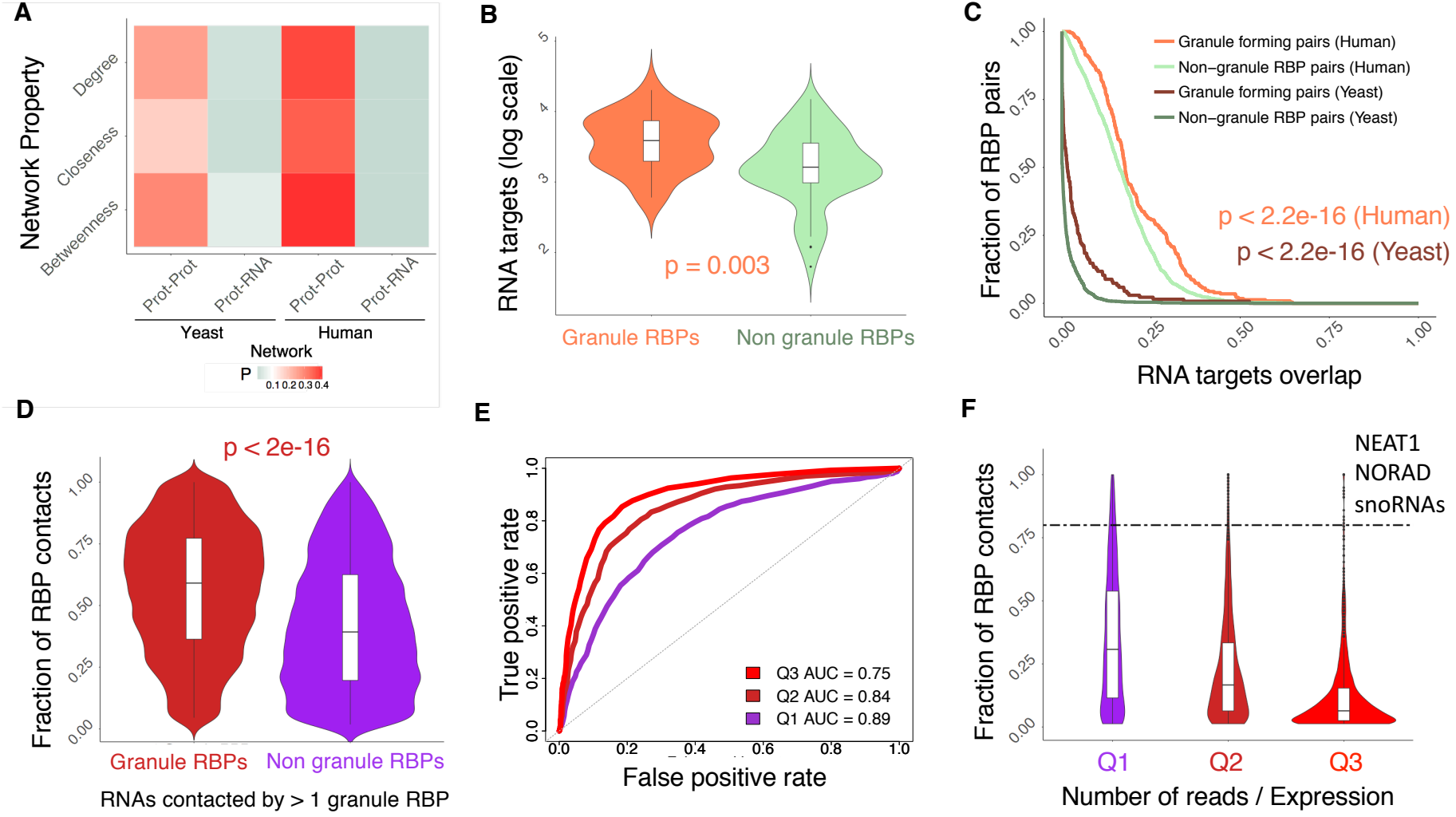
1. Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat. Struct. Mol. Biol.* **20,** 127–133 (2012).
2. Bolognesi, B. *et al.* A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. *Cell Rep.* **16,** 222–231 (2016).
3. Jain, S. *et al.* ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell* **164,** 487–498 (2016).
4. Buchan, J. R., Muhlrad, D. & Parker, R. P bodies promote stress granule assembly in Saccharomyces cerevisiae. *J. Cell Biol.* **183,** 441–455 (2008).
5. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30,** 39–58 (2014).
6. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149,** 753–767 (2012).
7. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).
8. Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88,** 678–690 (2015).
9. Pettersson, O. J., Aagaard, L., Jensen, T. G. & Damgaard, C. K. Molecular mechanisms in DM1 — a focus on foci. *Nucleic Acids Res.* **43,** 2433–2441 (2015).
10. Jonas, S. & Izaurralde, E. The role of disordered protein regions in the assembly of decapping complexes and RNP granules. *Genes Dev.* **27,** 2628–2641 (2013).
11. Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol.* **26,** 668–679 (2016).
12. Khong, A. *et al.* The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol. Cell* **0,** (2017).
13. Ribeiro, D. *et al.* Protein complex scaffolding predicted as a prevalent function of human long non-coding RNAs. *Nucleic Acids Research* (in press). doi:10.1093/nar/gkx1169
14. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13,** 508–514 (2016).

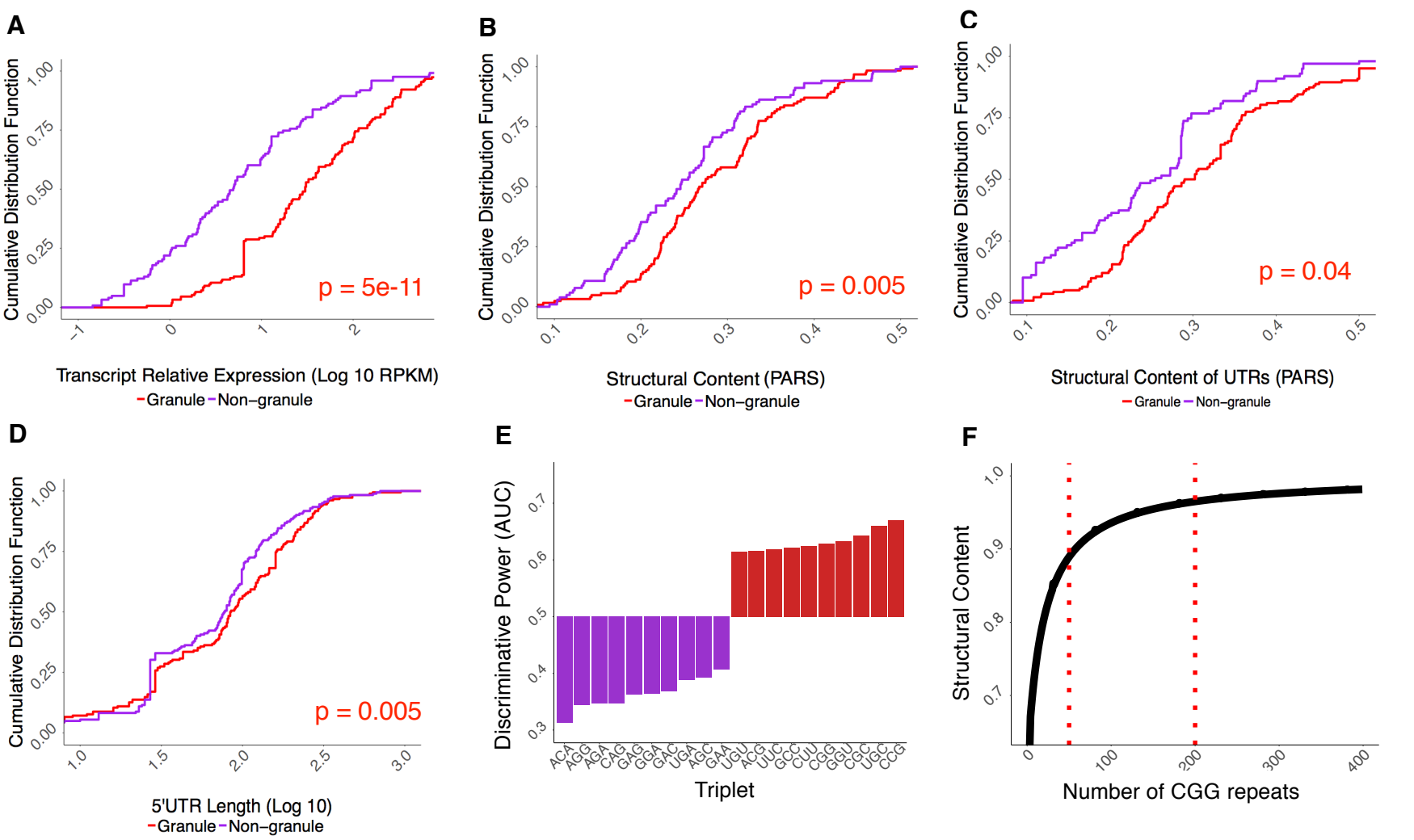
15. Brouwer, J. R., Willemsen, R. & Oostra, B. A. Microsatellite repeat instability and neurological disease. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **31,** 71–83 (2009).
16. Everett, C. M. & Wood, N. W. Trinucleotide repeats and neurodegenerative disease. *Brain J. Neurol.* **127,** 2385–2405 (2004).
17. Todd, P. K. *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78,** 440–455 (2013).
18. Botta-Orfila, T., Tartaglia, G. G. & Michalon, A. Molecular Pathophysiology of Fragile X-Associated Tremor/Ataxia Syndrome and Perspectives for Drug Development. *Cerebellum Lond. Engl.* **15,** 599–610 (2016).
19. Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain J. Neurol.* **129,** 256–271 (2006).
20. Sofola, O. A. *et al.* RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* **55,** 565–571 (2007).
21. Sellier, C. *et al*. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.* **29,** 1248–1261 (2010).
22. Sellier, C. *et al*. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep.* **3,** 869–880 (2013).
23. Qurashi, A., Li, W., Zhou, J.-Y., Peng, J. & Jin, P. Nuclear Accumulation of Stress Response mRNAs Contributes to the Neurodegeneration Caused by Fragile X Premutation rCGG Repeats. *PLOS Genet.* **7,** e1002102 (2011).
24. Sellier, C. *et al*. Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* **93,** 331–347 (2017).
25. Marchese, D., de Groot, N. S., Lorenzo Gotor, N., Livi, C. M. & Tartaglia, G. G. Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* **7,** 793–810 (2016).
26. Tartaglia, G. G. The Grand Challenge of Characterizing Ribonucleoprotein Networks. *Front. Mol. Biosci.* **3,** (2016).
27. Disney, M. D. *et al.* A Small Molecule that Targets r(CGG)exp and Improves Defects in Fragile X-Associated Tremor Ataxia Syndrome. *ACS Chem. Biol.* **7,** 1711–1718 (2012).

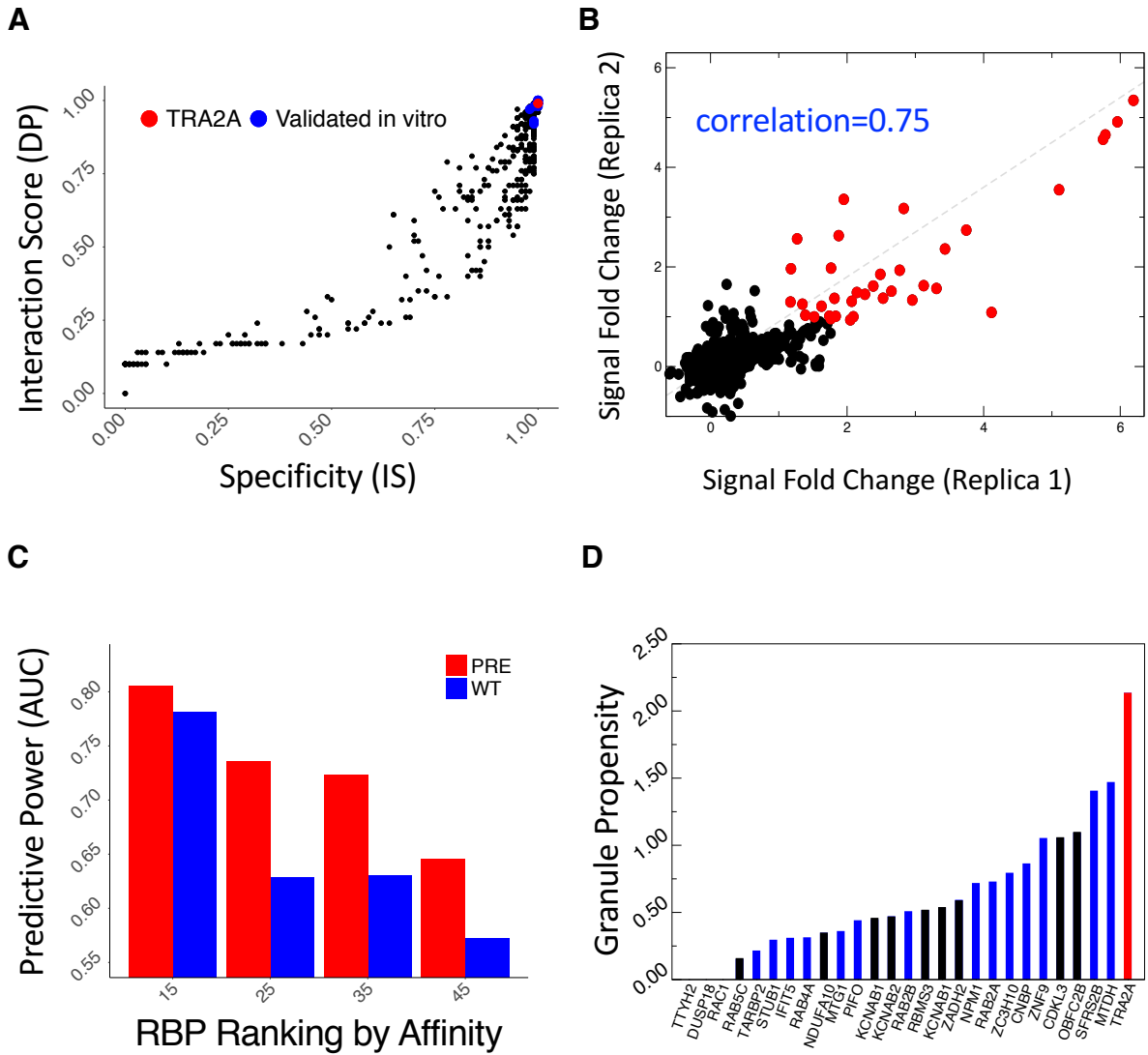
28. Marchese, D. *et al.* Discovering the 3′ UTR-mediated regulation of alpha-synuclein. *Nucleic Acids Res.* doi:10.1093/nar/gkx1048
29. Cirillo, D. *et al.* Quantitative predictions of protein interactions with long noncoding RNAs. *Nat. Methods* **14,** 5–6 (2017).
30. Mittal, N., Scherrer, T., Gerber, A. P. & Janga, S. C. Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. *J. Mol. Biol.* **409,** 466–479 (2011).
31. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162,** 425–440 (2015).
32. Armaos, A., Cirillo, D. & Tartaglia, G. G. omiXcore: a web server for prediction of protein interactions with large RNA. *Bioinforma. Oxf. Engl.* (2017). doi:10.1093/bioinformatics/btx361
33. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* **33,** 717–726 (2009).
34. Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164,** 69–80 (2016).
35. Tichon, A. *et al.* A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7,** ncomms12209 (2016).
36. Zhang, H. *et al.* RNA controls PolyQ protein phase transitions. *Mol. Cell* **60,** 220–230 (2015).
37. Reineke, L. C., Kedersha, N., Langereis, M. A., Kuppeveld, F. J. M. van & Lloyd, R. E. Stress Granules Regulate Double-Stranded RNA-Dependent Protein Kinase Activation through a Complex Containing G3BP1 and Caprin1. *mBio* **6,** e02486-14 (2015).
38. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546,** 243–247 (2017).
39. Krzyzosiak, W. J. *et al.* Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **40,** 11–26 (2012).
40. Delli Ponti, R., Marti, S., Armaos, A. & Tartaglia, G. G. A high-throughput approach to profile RNA structure. *Nucleic Acids Res.* **45,** e35–e35 (2017).
41. Agostini, F. *et al.* catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinforma. Oxf. Engl.* **29,** 2928–2930 (2013).

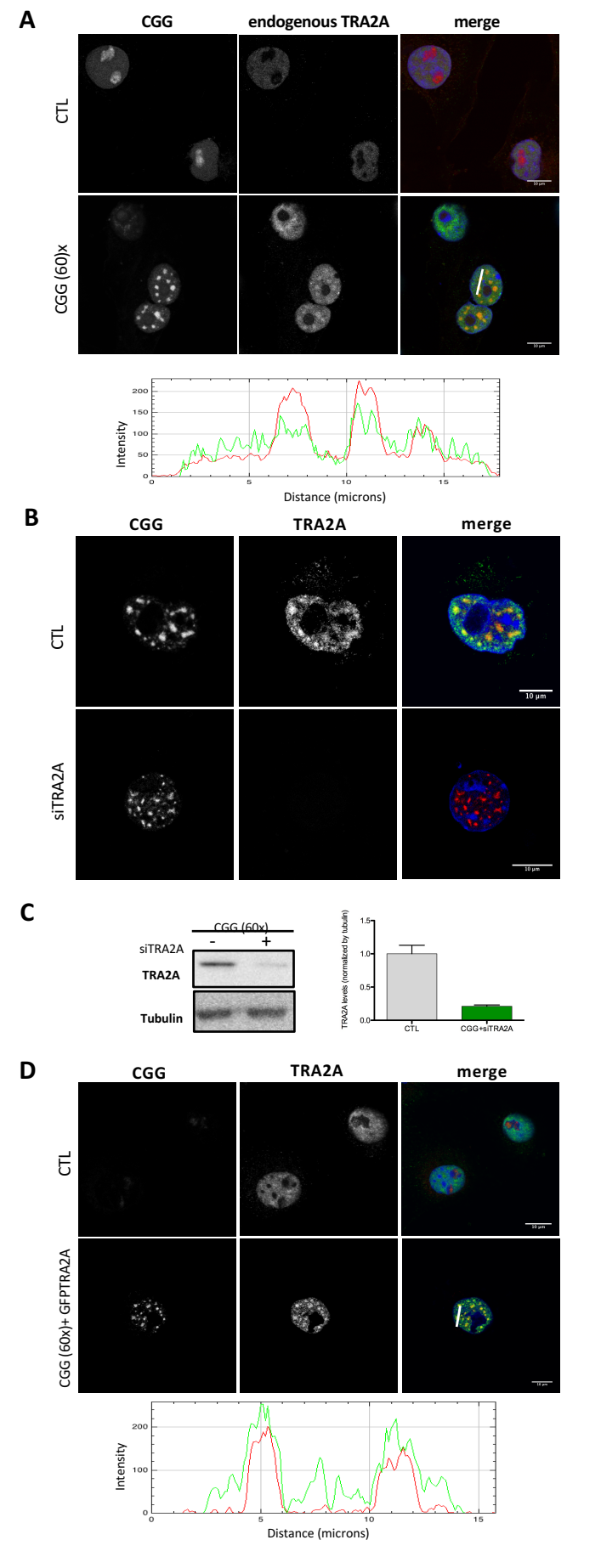
42. Livi, C. M., Klus, P., Delli Ponti, R. & Tartaglia, G. G. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv629
43. Cirillo, D. *et al.* Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. *RNA N. Y. N* **19,** 129–140 (2013).
44. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38,** W214–W220 (2010).
45. Klus, P., Ponti, R. D., Livi, C. M. & Tartaglia, G. G. Protein aggregation, structural disorder and RNA-binding ability: a new approach for physico-chemical and gene ontology classification of multiple datasets. *BMC Genomics* **16,** 1071 (2015).
46. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. 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,** 4137–4145 (2012).
47. Tyson-Capper, A. *et al.* Human Tra2 proteins jointly control a *CHEK1* splicing switch among alternative and constitutive target exons. *Nat. Commun.* **5,** ncomms5760 (2014).
48. Budny, B. *et al.* Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. *Clin. Genet.* **77,** 541–551 (2010).
49. Procaccio, V. *et al.* A mutation of beta -actin that alters depolymerization dynamics is associated with autosomal dominant developmental malformations, deafness, and dystonia. *Am. J. Hum. Genet.* **78,** 947–960 (2006).
50. Rivière, J.-B. *et al.* De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. *Nat. Genet.* **44,** 440–444, S1-2 (2012).
51. Narkis, G. *et al.* 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,** 530–539 (2007).
52. Erdmann, J. *et al.* Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. *Clin. Genet.* **64,** 339–349 (2003).
53. de Silva, M. G. *et al.* 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,** 733–740 (2003).

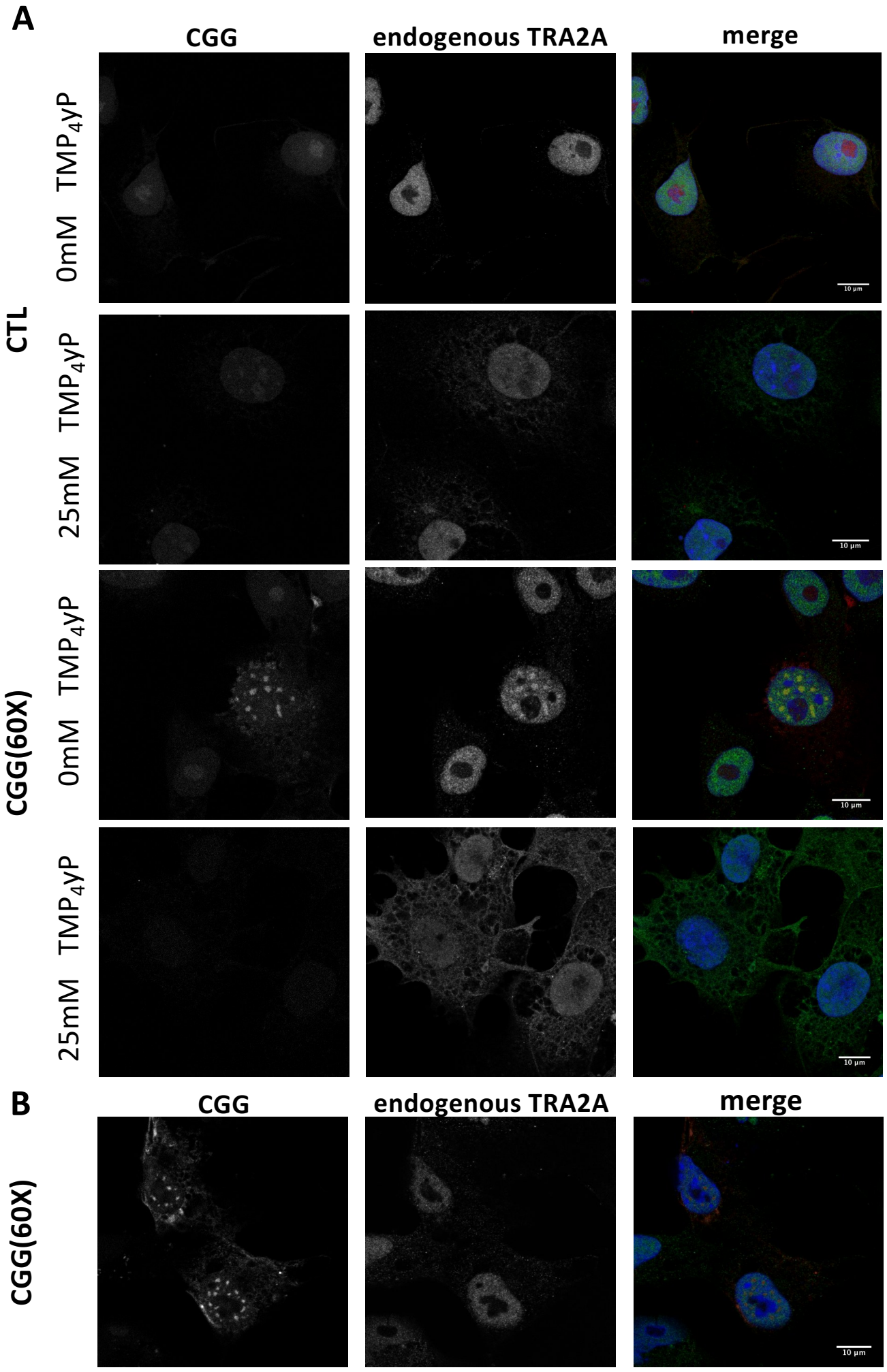
54. Buijsen, R. A. M. *et al.* FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol. Commun.* **2,** 162 (2014).
55. Hukema, R. K. *et al.* Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum. Mol. Genet.* **24,** 4948–4957 (2015).
56. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18,** 285–298 (2017).
57. Tacke, R., Tohyama, M., Ogawa, S. & Manley, J. L. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* **93,** 139–148 (1998).
58. Haynes, C. & Iakoucheva, L. M. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* **34,** 305–312 (2006).
59. Ying, Y. *et al.* Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* **170,** 312-323.e10 (2017).
60. Gueroussov, S. *et al.* Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* **170,** 324-339.e23 (2017).
61. Zlotorynski, E. Splicing: Phasing alternative exons. *Nat. Rev. Mol. Cell Biol.* **18,** 529 (2017).
62. Wood, J. D. *et al.* Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. *Mol. Cell. Neurosci.* **11,** 149–160 (1998).
63. Zou, W. *et al.* The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. *Nat. Cell Biol.* **13,** 59–65 (2011).
64. Brannan, K. W. *et al.* SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol. Cell* **64,** 282–293 (2016).
65. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15,** 829–845 (2014).

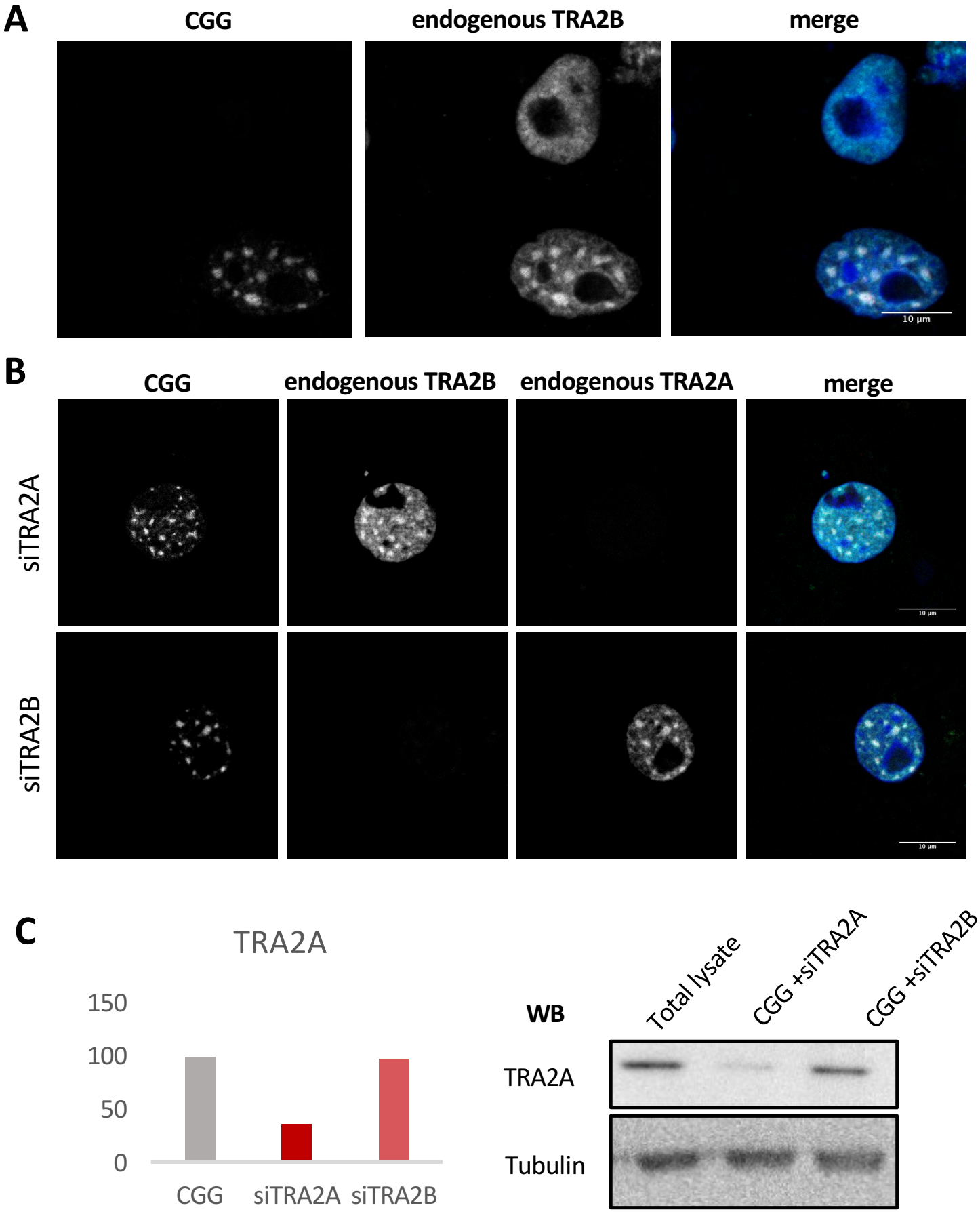

A
Network Property
Degree
Closeness
Betweenness
Prot-Prot
Prot-RNA
Prot-Prot
Prot-RNA
Yeast
Human
Network
P
0.1 0.2 0.3 0.4
B
RNA targets (log scale)
p = 0.003
Granule RBPs
Non granule RBPs
C
Fraction of RBP pairs
Granule forming pairs (Human)
Non–granule RBP pairs (Human)
Granule forming pairs (Yeast)
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
Q3 AUC = 0.75
Q2 AUC = 0.84
Q1 AUC = 0.89
False positive rate
F
NEAT1
NORAD
snoRNAs
Fraction of RBP contacts
Q1
Q2
Q3
Number of reads / Expression

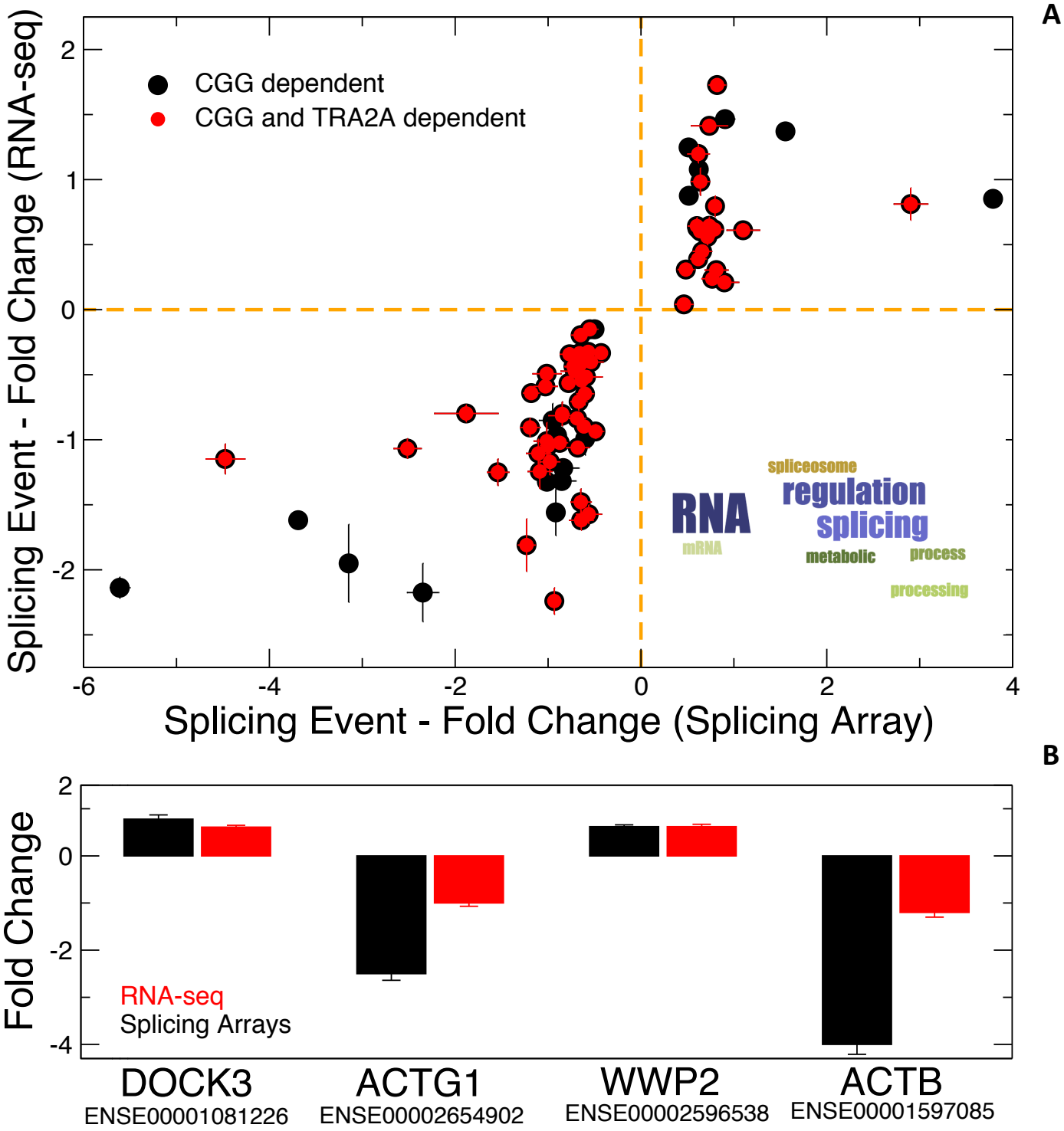

A
Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
−1
0
1
2
p = 5e-11
Transcript Relative Expression (Log 10 RPKM)
Granule
Non−granule
B
Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
0.1
0.2
0.3
0.4
0.5
p = 0.005
Structural Content (PARS)
Granule
Non−granule
C
Cumulative Distribution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

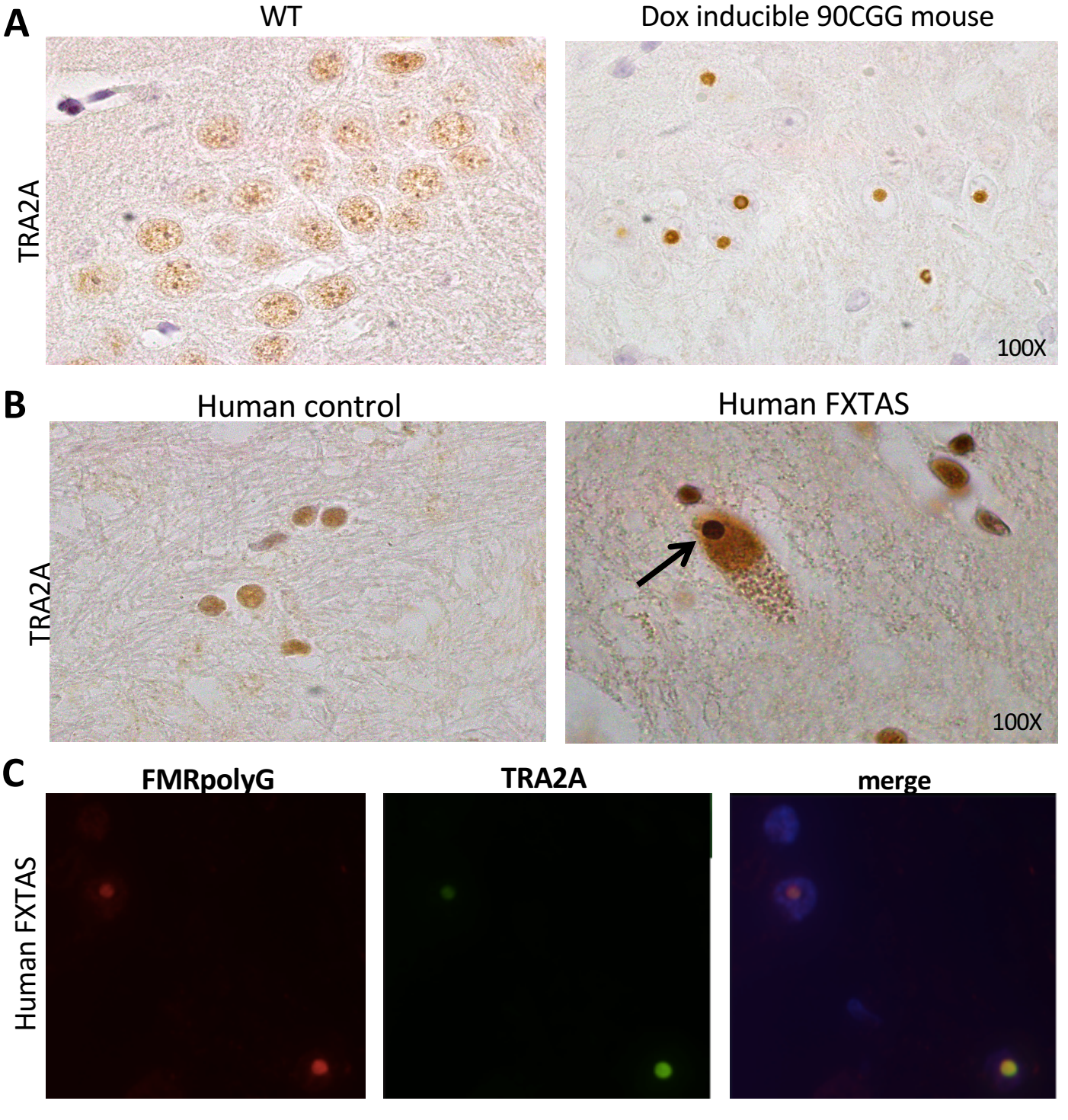

A
TRA2A
Validated in vitro
Interaction Score (DP)
Specificity (IS)
0.00
0.25
0.50
0.75
1.00
B
correlation=0.75
Signal Fold Change (Replica 2)
Signal Fold Change (Replica 1)
0
2
4
6
C
PRE
WT
Predictive Power (AUC)
RBP Ranking by Affinity
0.55
0.60
0.65
0.70
0.75
0.80
15
25
35
45
D
Granule Propensity
0.00
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


A
CGG
endogenous TRA2A
merge
CTL
CGG (60)x
Intensity
Distance (microns)
B
CGG
TRA2A
merge
CTL
siTRA2A
10 µm
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)


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


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


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
spliceosome
regulation
splicing
RNA
mRNA
metabolic
process
processing
B
Fold Change
2
0
-2
-4
RNA-seq
Splicing Arrays
DOCK3
ENSE00001081226
ACTG1
ENSE00002654902
WWP2
ENSE00002596538
ACTB
ENSE00001597085


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