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

Summary

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 phase-separating in RNP granules share specific RNA partners with high structural content, long untranslated regions (UTRs) and nucleotide repeat expansions. We predict that RNAs promote formation of RNP granules by acting as scaffolds for protein assembly.

To 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. We employed human primary tissues as well as primate and mouse models of FXTAS to characterise protein sequestration in FMR1 aggregates and identified a previously unknown partner, TRA2A, which induces impairment of splicing in genes linked to mental retardation.

Introduction

Ribonucleoprotein (RNP) aggregates or granules are liquid-like cellular compartments composed of RNA-binding proteins (RBPs) and transcript (Hyman et al., 2014; Maharana et al., 2018) that assemble through a process of phase separation and are in dynamic exchange with the surrounding environment (Bolognesi et al., 2016). RNP granules such as processing bodies and stress granules (SGs) are evolutionarily conserved from yeast to human (Jain et al., 2016) and contain constitutive protein components such as G3BP1 (yeast: Nxt3), TIA1 (Pub1), and TIAR (Ngr1) (Buchan et al., 2008). Several RBPs involved in physiological liquid-liquid phase separations are prone to form amyloid aggregates upon amino acid mutations (Hyman et al., 2014; Kato et al., 2012) that induce a transition from a liquid droplet to a solid phase (Qamar et al., 2018). This observation has led to the proposal that a liquid-to-solid phase transition is a mechanism of cellular toxicity (Patel et al., 2015). in diseases such as Amyotrophic Lateral Sclerosis (Murakami et al., 2015) and Myotonic Dystrophy (Pettersson et al., 2015).

Recent evidence indicates that RBPs act as scaffolding elements promoting RNP granule assembly through protein-protein interactions (PPI) (Jonas and Izaurralde, 2013; Protter and Parker, 2016), 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 (Khong et al., 2017). 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 (Ribeiro et al., 2018), 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 PPI and PRI networks revealed by enhanced CrossLinking and ImmunoPrecipitation (eCLIP) (Van Nostrand et al., 2016), mass spectrometric analysis of SGs (Jain et al., 2016) and novel computational methods (Delli Ponti et al., 2017), 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 (Berkovits and Mayr, 2015), 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 (Saha and Hyman, 2017). In support of this observation, it has been reported for several diseases including Huntington's disease, Myotonic Dystrophy and a number of Ataxias that expanded trinucleotide regions trigger formation of intranuclear aggregates in which proteins are sequestered and inactivated (Brouwer et al., 2009; Everett and Wood, 2004).

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 (Todd et al., 2013). 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) (Todd et al., 2013). These *foci* are highly dynamic and behave as ribonucleoprotein granules (Strack et al., 2013) that phase separate in the nucleus forming inclusions (Tassone et al., 2004). Importantly, CGG aggregates are long lived bur rapidly dissolve upon tautomycin treatment, which indicates liquid-like aggregation (Strack and Jaffrey, 2014).

Two mechanisms have been suggested to explain onset and development of FXTAS (Botta-Orfila et al., 2016): 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) (Todd et al., 2013). Previous work indicates that *FMR1* inclusions contain specific proteins such as HNRNP A2/B1, MBNL1, LMNA and INA (Iwahashi et al., 2006). Also FMRpolyG peptides (Sellier et al., 2017) have been found in the inclusions, together with CUGBP1, KHDRBS1 and DGCR8 that are involved in splicing regulation, mRNA transport and regulation of microRNAs (Qurashi et al., 2011; Sellier et al., 2010, 2013; Sofola et al., 2007). Importantly, KHDRBS1 does not directly bind to *FMR1* (Sellier et al., 2010), while its protein partner DGCR8 interacts physically with CGG repeats (Sellier

et al., 2013), 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 (Marchese et al., 2016; Tartaglia, 2016), 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 (Sellier et al., 2017; Todd et al., 2013) and a strong demand for new approaches (Disney et al., 2012).

Results

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 (**Supplemental Information** and **Supplementary Table 1A**) as well as novel computational methods. The protein-RNA interaction datasets were identified through eCLIP (human) (Van Nostrand et al., 2016) and microarray (yeast) (Mittal et al., 2011) 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 **(Supplemental Information)** (Huttlin et al., 2015; Mittal et al., 2011). 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; **Supplemental Information**). 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 (Armaos et al., 2017)), 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**) (Khong et al., 2017).

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 (Maharana et al., 2018) for paraspeckle nucleation (Clemson et al., 2009). 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 (Lee et al., 2016). NORAD has repetitive sequence units, activation upon stress and ability to sequester proteins (Tichon et al., 2016), 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 (**Supplemental Information**), we found that RNAs enriched in granule RBP contacts are more expressed (**Figure 2A**; p-value = 5e-11, KS test; **Supplemental Information**), 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 (Zhang et al., 2015), structure (Reineke et al., 2015) and abundance (Jain and Vale, 2017) contribute to RNA assembly into RNP granules (Khong et al., 2017).

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 (Krzyzosiak et al., 2012): 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 (Delli Ponti et al., 2017) 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 (Krzyzosiak et al., 2012).

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* (**Supplemental Information**) (Agostini et al., 2013), 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 (Livi et al., 2015)). Previously identified CGG-binding proteins (Sellier et al., 2010) 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 (Jain et al., 2016) (**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 (Bolognesi et al., 2016) (**Supplemental Information; 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 (Cirillo et al., 2017; Marchese et al., 2017). 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) (Cirillo et al., 2013, 2017).

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 (Warde-Farley et al., 2010) and *clever*GO (Klus et al., 2015)) 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 (Cirillo et al., 2013), and DGCR8, which interacts with KHDRBS1 through DROSHA (Sellier et al., 2013), is also prone to interact (top 7% of all fluorescence intensities).

Out of 27 high-confidence candidates, 24 were predicted by *cat*GRANULE (Bolognesi et al., 2016) 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 (Van Nostrand et al., 2016).

TRA2A recruitment in FMR1 inclusions is driven by CGG hairpins

As splicing defects have been reported to occur in FXTAS disease (Botta-Orfila et al., 2016; Sellier et al., 2010), 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 (Sellier et al., 2010). 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**) (Sellier et al., 2010). By means of RNA FISH coupled to immunofluorescence (**Supplemental Information**), we found that CGG expansions form nuclear inclusions that co-localize with endogenous TRA2A (**Supplemental Information**). TRA2A shows a diffuse nuclear pattern in cells that do not over-express CGG repeats (**Figure 4A**).

Upon knockdown of TRA2A using siRNA (**Supplemental Information**) 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 in 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 (Disney et al., 2012), TRA2A sequestration was blocked (**Figure 5A**). Using TmPyP4 to specifically unfold CGG repeats (Morris et al., 2012) , 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**) (Qurashi et al., 2011; Sellier et al., 2010, 2013; Sofola et al., 2007). 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 (**Supplemental Information**).

Splicing alterations due to TRA2A sequestration were studied through microarrays and RNA-seq experiments (both in triplicate experiments; **Supplemental Information**), 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; Supplemental Information**). Using the *clever*GO algorithm (Klus et al., 2015) 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: HNRNP A2/B1 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 (Budny et al., 2010), ACTB (Procaccio et al., 2006) and ACTG1 (Rivière et al., 2012), have splicing patterns affected by TRA2A sequestration. Similarly, muscle related proteins, including PIP5K1C (Narkis et al., 2007), TPM1 (Erdmann et al., 2003) and genes linked to intellectual disabilities such as DOCK3 (de Silva et al., 2003) and craniofacial development, such as WWP (Wood et al., 1998; Zou et al., 2011), 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 (Maurin et al., 2014).

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 (Todd et al., 2013). The main RAN translation product, FMRpolyG, co-localizes with ubiquitin in intranuclear inclusions, as observed in FXTAS human brain and mouse models (Buijsen et al., 2014; Sellier et al., 2017; Todd et al., 2013).

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 (Hukema et al., 2015). Immunohistochemistry experiments with sections of paraffin-embedded neurons and astrocytes show that TRA2A protein is indeed present in the inclusions (**Figure 7C**).

Remarkably, nuclear inclusions from FXTAS *post mortem* human brain donors show positive staining for TRA2A (**Figure 7D**) that also co-localize with FMRpolyG (**Figure 7E**).

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.

Discussion

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 (Banani et al., 2017). 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 (Botta-Orfila et al., 2016; Sellier et al., 2010). Combining *in silico* predictions of protein-RNA interactions with large-scale *in vitro* validations (Cirillo et al., 2017; Marchese et al., 2017), 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 (Sellier et al., 2010), 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* (Sellier et al., 2013).

At the time of writing, TRA2A has been reported to be a constitutive component of granules associated with Amyotrophic Lateral Sclerosis (Markmiller et al., 2018), 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* (Tacke et al., 1998). Similarly to other splicing factors, TRA2A contains structurally disordered regions and multiple stretches of low

sequence complexity (Haynes and Iakoucheva, 2006) that promote liquid-liquid phase separation (Bolognesi et al., 2016). Although formation of large assemblies is crucial for splicing activity and regulation (Gueroussov et al., 2017; Ying et al., 2017), 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 (Pettersson et al., 2015) leading to impairment of splicing activity and Myotonic Dystrophy (Zlotorynski, 2017).

In a cellular model of FXTAS, we found that TRA2A is sequestered by nuclear RNA inclusions upon over-expression of *FMR1* 5' UTR. In agreement with our computational analysis, PPI networks do not play a main role in protein recruitment, as identified in our *in silico* analysis: both endogenous and overexpressed TRA2A are recruited by CGG granules and upon knockdown of TRA2B (Tyson-Capper et al., 2014) the assembly process is still driven by RNA. Unfolding the structure of CGG repeats through TMP4yP (Morris et al., 2012) 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 (Disney et al., 2012) results in preventing TRA2A sequestration. Through splicing microarrays and RNA-seq analysis we found that TRA2A sequestration induces changes in 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 very relevant in the context of Fragile X Syndrome (Bardoni et al., 2006).

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* (Todd et al., 2013) 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 (Hlevnjak et al., 2012; Zanzoni et al., 2013).

Our work highlights general features that characterize scaffolding RNAs 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 (Jain and Vale, 2017). Moreover, two articles published at the time of writing reported that the structural content (Maharana et al., 2018) and UTR length (Khong et al., 2017) 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 (Knowles et al., 2009; Tartaglia and Vendruscolo, 2009; Speretta et al., 2012) but also applies to RNA scaffolds.

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 all members of Tartaglia's lab and especially Dr. Elias Bechara for the designing and interpretation of splicing experiments, Dr Lara Nonell Mazelón and Dr Magdalena Arnal Segura for RNAseq and splicing arrays analysis, 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 and TBO conceived the study together with BB, FCS performed the calculations, TBO, IGB, RH, LWS and EG performed the experiments. TBO, NLG, GGT and BL analyzed the data. TBO, FCS, 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 (Jain et al., 2016) (Area under the ROC curve AUC is used to measure the enrichment in SGs; **Supplemental Information**). 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 (Delli Ponti et al., 2017) is employed to measure the amount of double-stranded RNA).

Figure 3. Protein interactions of CGG repeats. A) Using *cat*RAPID *omics* (Agostini et al., 2013), we computed protein interactions with the first *FMR1* exon (containing 79 CGG repeats). Previously identified partners, such as HNRNP A1,

A2/B1, A3, 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) (Sellier et al., 2010). 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 (Cirillo et al., 2017; Marchese et al., 2017). 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* (Agostini et al., 2013) 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) (Bolognesi et al., 2016) and the splicing regulator TRA2A (red bar) shows the highest propensity. The black bars indicate non-specific partners (interacting also with SNCA 3' UTR (Cirillo et al., 2017; Marchese et al., 2017) or showing poor RNA-binding propensity (Livi et al., 2015)).

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 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; Supplemental Information**). **Inset**: GO analysis of the largest cluster (18 genes) affected by TRA2A sequestration (Klus et al., 2015) 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**). **TRA2A accumulations are present in murine and human FXTAS inclusions. C)** TRA2A immunohistochemistry in wild type (WT) and premutated mouse model (counterstaining is done with haematoxylin). **D)** TRA2A immunohistochemistry in human hippocampus from control and FXTAS (counterstained with haematoxylin; the arrows points to the inclusion). **E)** Double immunofluorescence of TRA2A as well as FMRpolyG peptides in human FXTAS (**Supplemental Information**).

Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D., and Tartaglia, G.G. (2013). catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. Bioinforma. Oxf. Engl. *29*, 2928–2930.

Armaos, A., Cirillo, D., and Tartaglia, G.G. (2017). omiXcore: a web server for prediction of protein interactions with large RNA. Bioinforma. Oxf. Engl.

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

Bardoni, B., Davidovic, L., Bensaid, M., and Khandjian, E.W. (2006). The fragile X syndrome: exploring its molecular basis and seeking a treatment. Expert Rev. Mol. Med. *8*, 1–16.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363–367.

Bolognesi, B., Lorenzo Gotor, N., Dhar, R., Cirillo, D., Baldrighi, M., Tartaglia, G.G., and Lehner, B. (2016). A Concentration-Dependent Liquid Phase Separation Can Cause Toxicity upon Increased Protein Expression. Cell Rep. *16*, 222–231.

Botta-Orfila, T., Tartaglia, G.G., and Michalon, A. (2016). Molecular Pathophysiology of Fragile X-Associated Tremor/Ataxia Syndrome and Perspectives for Drug Development. Cerebellum Lond. Engl. *15*, 599–610.

Brouwer, J.R., Willemsen, R., and Oostra, B.A. (2009). Microsatellite repeat instability and neurological disease. BioEssays News Rev. Mol. Cell. Dev. Biol. *31*, 71–83.

Buchan, J.R., Muhlrad, D., and Parker, R. (2008). P bodies promote stress granule assembly in Saccharomyces cerevisiae. J. Cell Biol. *183*, 441–455.

Budny, B., Badura-Stronka, M., Materna-Kiryluk, A., Tzschach, A., Raynaud, M., Latos-Bielenska, A., and Ropers, H.H. (2010). Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. Clin. Genet. *77*, 541–551.

Buijsen, R.A.M., Sellier, C., Severijnen, L.-A.W.F.M., Oulad-Abdelghani, M., Verhagen, R.F.M., Berman, R.F., Charlet-Berguerand, N., Willemsen, R., and Hukema, R.K. (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.

Cirillo, D., Agostini, F., Klus, P., Marchese, D., Rodriguez, S., Bolognesi, B., and Tartaglia, G.G. (2013). Neurodegenerative diseases: quantitative predictions of protein-RNA interactions. RNA N. Y. N *19*, 129–140.

Cirillo, D., Blanco, M., Armaos, A., Buness, A., Avner, P., Guttman, M., Cerase, A., and Tartaglia, G.G. (2017). Quantitative predictions of protein interactions with long noncoding RNAs. Nat. Methods *14*, 5–6.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. Mol. Cell *33*, 717–726.

Delli Ponti, R., Marti, S., Armaos, A., and Tartaglia, G.G. (2017). A high-throughput approach to profile RNA structure. Nucleic Acids Res. *45*, e35–e35.

Disney, M.D., Liu, B., Yang, W.-Y., Sellier, C., Tran, T., Charlet-Berguerand, N., and Childs-Disney, J.L. (2012). A Small Molecule that Targets r(CGG)exp and Improves Defects in Fragile X-Associated Tremor Ataxia Syndrome. ACS Chem. Biol. *7*, 1711–1718.

Erdmann, J., Daehmlow, S., Wischke, S., Senyuva, M., Werner, U., Raible, J., Tanis, N., Dyachenko, S., Hummel, M., Hetzer, R., et al. (2003). Mutation spectrum in a large cohort of unrelated consecutive patients with hypertrophic cardiomyopathy. Clin. Genet. *64*, 339–349.

Everett, C.M., and Wood, N.W. (2004). Trinucleotide repeats and neurodegenerative disease. Brain J. Neurol. *127*, 2385–2405.

Gueroussov, S., Weatheritt, R.J., O'Hanlon, D., Lin, Z.-Y., Narula, A., Gingras, A.-C., and Blencowe, B.J. (2017). Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. Cell *170*, 324-339.e23.

Haynes, C., and Iakoucheva, L.M. (2006). Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. Nucleic Acids Res. *34*, 305–312.

Hlevnjak, M., Polyansky, A.A., and Zagrovic, B. (2012). Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. Nucleic Acids Res. *40*, 8874–8882.

Hukema, R.K., Buijsen, R.A.M., Schonewille, M., Raske, C., Severijnen, L.-A.W.F.M., Nieuwenhuizen-Bakker, I., Verhagen, R.F.M., van Dessel, L., Maas, A., Charlet-Berguerand, N., et al. (2015). Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. Hum. Mol. Genet. *24*, 4948–4957.

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

Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-Liquid Phase Separation in Biology. Annu. Rev. Cell Dev. Biol. *30*, 39–58.

Iwahashi, C.K., Yasui, D.H., An, H.-J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J., and Hagerman, P.J. (2006). Protein composition of the intranuclear inclusions of FXTAS. Brain J. Neurol. *129*, 256–271.

Jain, A., and Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature *546*, 243–247.

Jain, S., Wheeler, J.R., Walters, R.W., Agrawal, A., Barsic, A., and Parker, R. (2016). ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. Cell *164*, 487–498.

Jonas, S., and Izaurralde, E. (2013). The role of disordered protein regions in the assembly of decapping complexes and RNP granules. Genes Dev. *27*, 2628–2641.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Khong, A., Matheny, T., Jain, S., Mitchell, S.F., Wheeler, J.R., and Parker, R. (2017). The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. Mol. Cell *0*.

Klus, P., Ponti, R.D., Livi, C.M., and Tartaglia, G.G. (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*, 1071.

Knowles, T.P.J., Waudby, C.A., Devlin, G.L., Cohen, S.I.A., Aguzzi, A., Vendruscolo, M., Terentjev, E.M., Welland, M.E., and Dobson, C.M. (2009). An Analytical Solution to the Kinetics of Breakable Filament Assembly. Science *326*, 1533–1537.

Krzyzosiak, W.J., Sobczak, K., Wojciechowska, M., Fiszer, A., Mykowska, A., and Kozlowski, P. (2012). Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. Nucleic Acids Res. *40*, 11–26.

Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell *164*, 69–80.

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., and 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*, D92–D97.

Livi, C.M., Klus, P., Delli Ponti, R., and Tartaglia, G.G. (2015). catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. Bioinforma. Oxf. Engl.

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

Marchese, D., de Groot, N.S., Lorenzo Gotor, N., Livi, C.M., and Tartaglia, G.G. (2016). Advances in the characterization of RNA-binding proteins. Wiley Interdiscip. Rev. RNA *7*, 793–810.

Marchese, D., Botta-Orfila, T., Cirillo, D., Rodriguez, J.A., Livi, C.M., Fernández-Santiago, R., Ezquerra, M., Martí, M.J., Bechara, E., Tartaglia, G.G., et al. (2017). Discovering the 3′ UTR-mediated regulation of alpha-synuclein. Nucleic Acids Res. *45*, 12888–12903.

Markmiller, S., Soltanieh, S., Server, K.L., Mak, R., Jin, W., Fang, M.Y., Luo, E.-C., Krach, F., Yang, D., Sen, A., et al. (2018). Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. Cell *172*, 590-604.e13.

Maurin, T., Zongaro, S., and Bardoni, B. (2014). Fragile X Syndrome: from molecular pathology to therapy. Neurosci. Biobehav. Rev. *46 Pt 2*, 242–255.

Mittal, N., Scherrer, T., Gerber, A.P., and Janga, S.C. (2011). Interplay between posttranscriptional and posttranslational interactions of RNA-binding proteins. J. Mol. Biol. *409*, 466–479.

Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C., and 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*, 4137–4145.

Murakami, T., Qamar, S., Lin, J.Q., Schierle, G.S.K., Rees, E., Miyashita, A., Costa, A.R., Dodd, R.B., Chan, F.T.S., Michel, C.H., 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*, 678–690.

Narkis, G., Ofir, R., Landau, D., Manor, E., Volokita, M., Hershkowitz, R., Elbedour, K., and Birk, O.S. (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*, 530–539.

Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066–1077.

Pettersson, O.J., Aagaard, L., Jensen, T.G., and Damgaard, C.K. (2015). Molecular mechanisms in DM1 — a focus on foci. Nucleic Acids Res. *43*, 2433–2441.

Procaccio, V., Salazar, G., Ono, S., Styers, M.L., Gearing, M., Davila, A., Jimenez, R., Juncos, J., Gutekunst, C.-A., Meroni, G., 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*, 947–960.

Protter, D.S.W., and Parker, R. (2016). Principles and Properties of Stress Granules. Trends Cell Biol. *26*, 668–679.

Qamar, S., Wang, G., Randle, S.J., Ruggeri, F.S., Varela, J.A., Lin, J.Q., Phillips, E.C., Miyashita, A., Williams, D., Ströhl, F., et al. (2018). FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. Cell *173*, 720-734.e15.

Qurashi, A., Li, W., Zhou, J.-Y., Peng, J., and Jin, P. (2011). Nuclear Accumulation of Stress Response mRNAs Contributes to the Neurodegeneration Caused by Fragile X Premutation rCGG Repeats. PLOS Genet. *7*, e1002102.

Reineke, L.C., Kedersha, N., Langereis, M.A., Kuppeveld, F.J.M. van, and Lloyd, R.E. (2015). Stress Granules Regulate Double-Stranded RNA-Dependent Protein Kinase Activation through a Complex Containing G3BP1 and Caprin1. MBio *6*, e02486-14.

Ribeiro, D.M., Zanzoni, A., Cipriano, A., Delli Ponti, R., Spinelli, L., Ballarino, M., Bozzoni, I., Tartaglia, G.G., and Brun, C. (2018). Protein complex scaffolding predicted as a prevalent function of long non-coding RNAs. Nucleic Acids Res. *46*, 917–928.

Rivière, J.-B., van Bon, B.W.M., Hoischen, A., Kholmanskikh, S.S., O'Roak, B.J., Gilissen, C., Gijsen, S., Sullivan, C.T., Christian, S.L., Abdul-Rahman, O.A., et al. (2012). De novo mutations in the actin genes ACTB and ACTG1 cause Baraitser-Winter syndrome. Nat. Genet. *44*, 440–444, S1-2.

Saha, S., and Hyman, A.A. (2017). RNA gets in phase. J. Cell Biol. *216*, 2235–2237.

Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J., et al. (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J. *29*, 1248–1261.

Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., 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*, 869–880.

Sellier, C., Buijsen, R.A.M., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A., et al. (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron *93*, 331–347.

de Silva, M.G., Elliott, K., Dahl, H.-H., Fitzpatrick, E., Wilcox, S., Delatycki, M., Williamson, R., Efron, D., Lynch, M., and Forrest, S. (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*, 733–740.

Sofola, O.A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D.L., and Botas, J. (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*, 565–571.

Speretta, E., Jahn, T.R., Tartaglia, G.G., Favrin, G., Barros, T.P., Imarisio, S., Lomas, D.A., Luheshi, L.M., Crowther, D.C., and Dobson, C.M. (2012). Expression in Drosophila of Tandem Amyloid β Peptides Provides Insights into Links between Aggregation and Neurotoxicity. J. Biol. Chem. *287*, 20748–20754.

Strack, R.L., and Jaffrey, S.R. (2014). Chapter 6 - Using RNA Mimics of GFP to Image RNA Dynamics in Mammalian Cells. In Fluorescence Microscopy, (Boston: Academic Press), pp. 83–91.

Strack, R.L., Disney, M.D., and Jaffrey, S.R. (2013). A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. Nat. Methods *10*, 1219–1224.

Tacke, R., Tohyama, M., Ogawa, S., and Manley, J.L. (1998). Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. Cell *93*, 139–148.

Tartaglia, G.G. (2016). The Grand Challenge of Characterizing Ribonucleoprotein Networks. Front. Mol. Biosci. *3*.

Tartaglia, G.G., and Vendruscolo, M. (2009). Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. Mol. Biosyst. *5*, 1873–1876.

Tassone, F., Iwahashi, C., and Hagerman, P.J. (2004). FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). RNA Biol. *1*, 103–105.

Tichon, A., Gil, N., Lubelsky, Y., Solomon, T.H., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nat. Commun. *7*, ncomms12209.

Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K., Scaglione, K.M., Basrur, V., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440–455.

Tyson-Capper, A., Best, A., Wipat, A., Chabot, B., Keavney, B., Austin, C.A., Dalgliesh, C., Elliott, D.J., Hong, E., Cowell, I.G., et al. (2014). Human Tra2 proteins jointly control a *CHEK1* splicing switch among alternative and constitutive target exons. Nat. Commun. *5*, ncomms5760.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. *38*, W214–W220.

Wood, J.D., Yuan, J., Margolis, R.L., Colomer, V., Duan, K., Kushi, J., Kaminsky, Z., Kleiderlein, J.J., Sharp, A.H., and Ross, C.A. (1998). Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. Mol. Cell. Neurosci. *11*, 149–160.

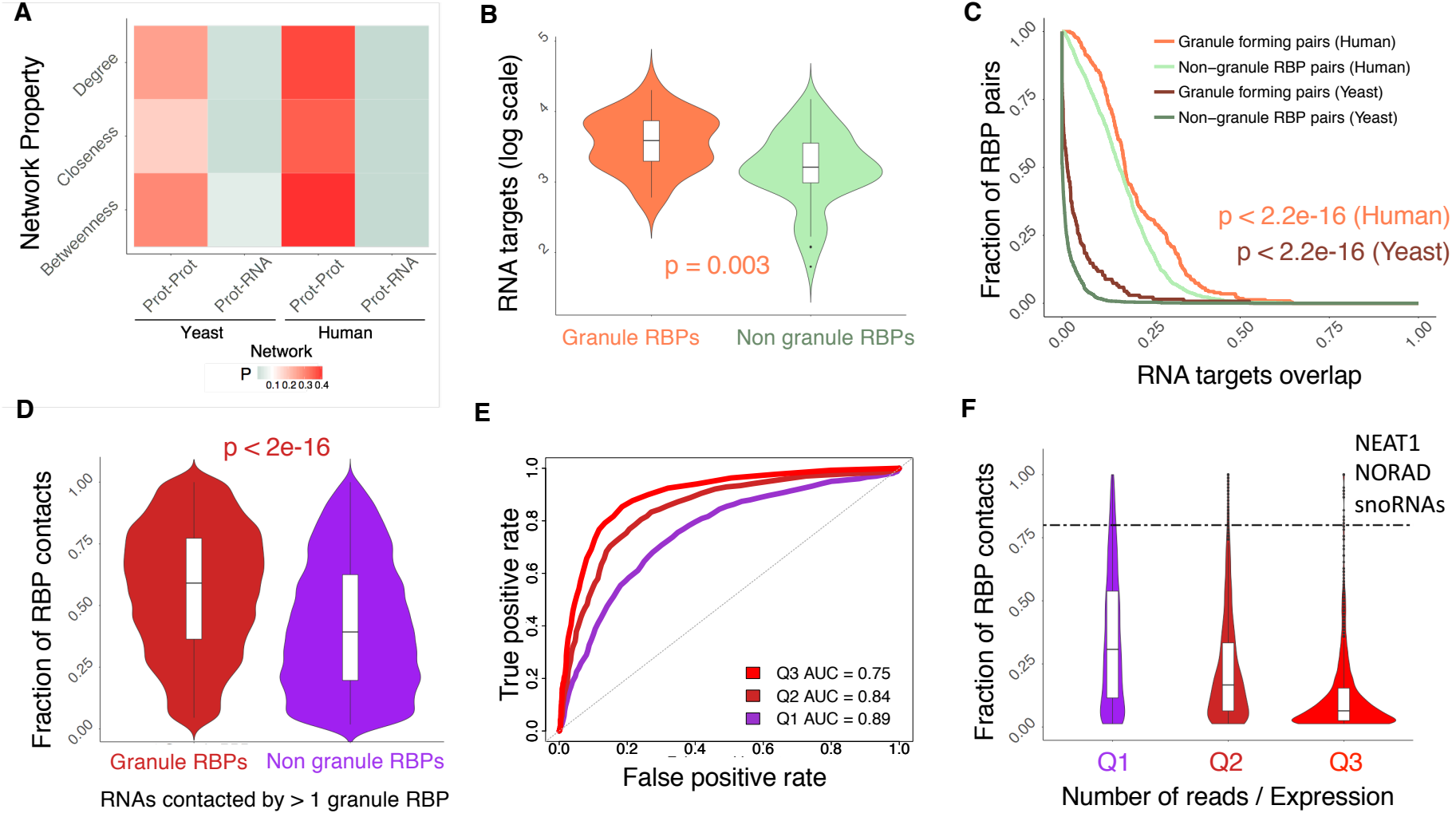
Ying, Y., Wang, X.-J., Vuong, C.K., Lin, C.-H., Damianov, A., and Black, D.L. (2017). Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. Cell *170*, 312-323.e10.

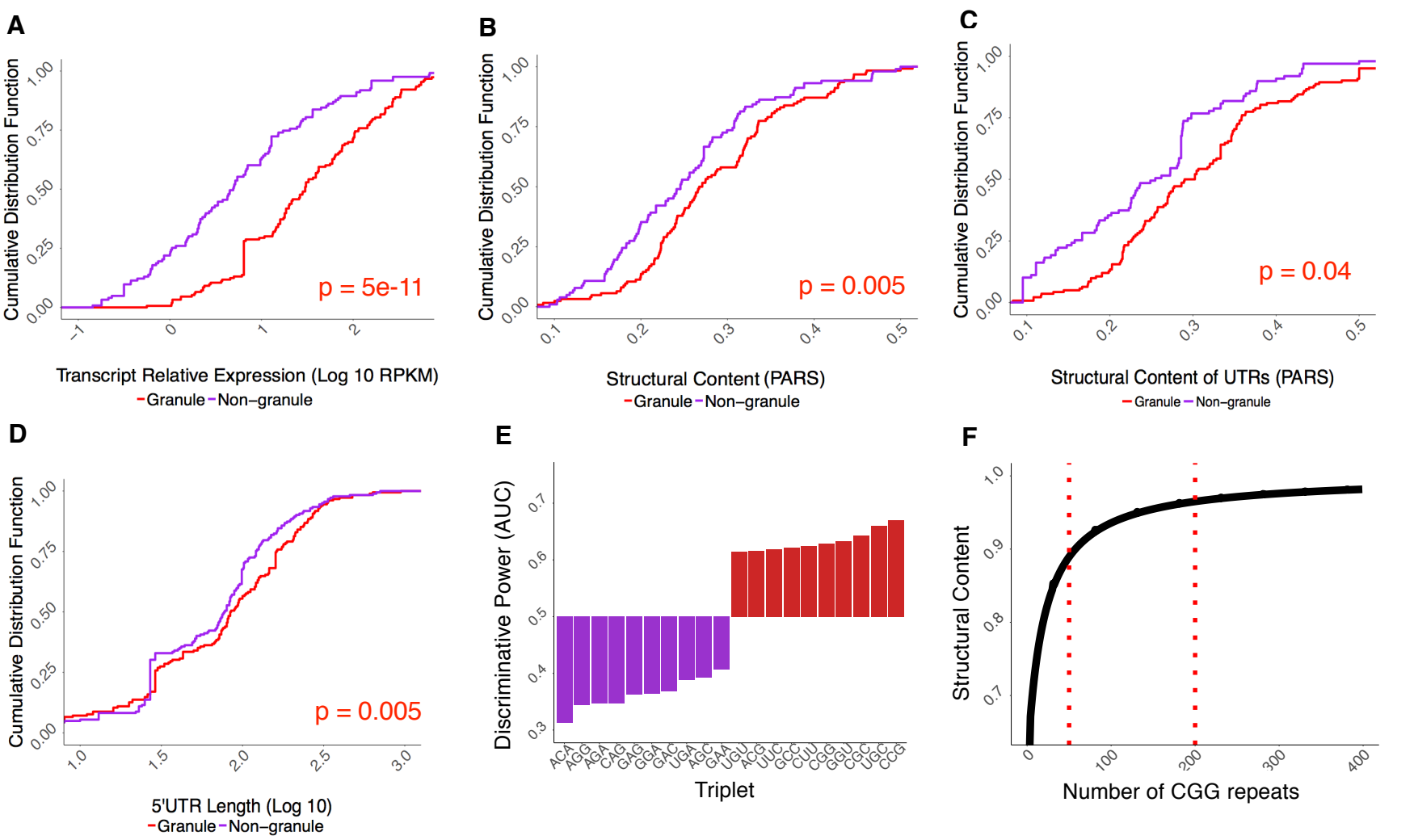
Zanzoni, A., Marchese, D., Agostini, F., Bolognesi, B., Cirillo, D., Botta-Orfila, M., Livi, C.M., Rodriguez-Mulero, S., and Tartaglia, G.G. (2013). Principles of self-organization in biological pathways: a hypothesis on the autogenous association of alpha-synuclein. Nucleic Acids Res. *41*, 9987–9998.

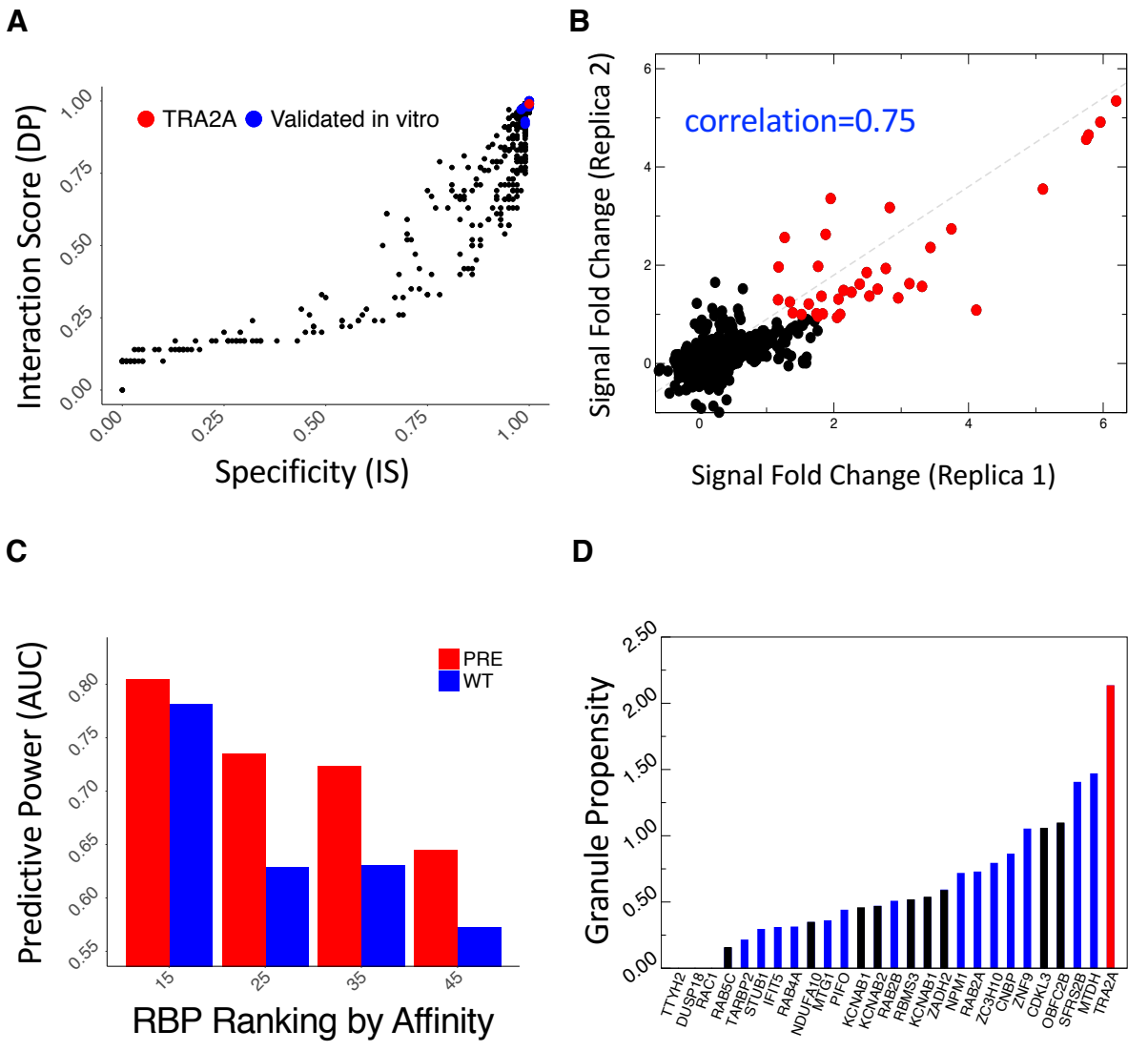
Zhang, H., Elbaum-Garfinkle, S., Langdon, E., Taylor, N., Occhipinti, P., Bridges, A., Brangwynne, C.P., and Gladfelter, A.S. (2015). RNA controls PolyQ protein phase transitions. Mol. Cell *60*, 220–230.

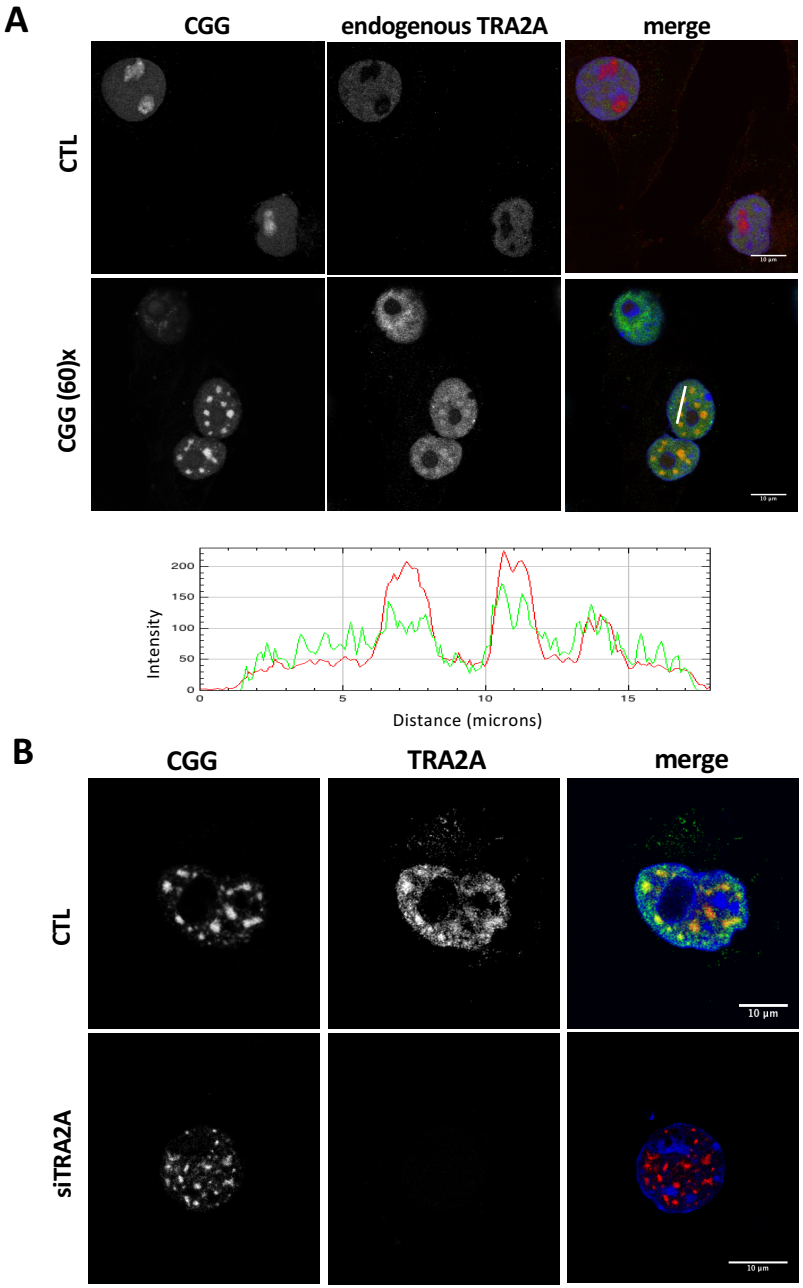
Zlotorynski, E. (2017). Splicing: Phasing alternative exons. Nat. Rev. Mol. Cell Biol. *18*, 529.

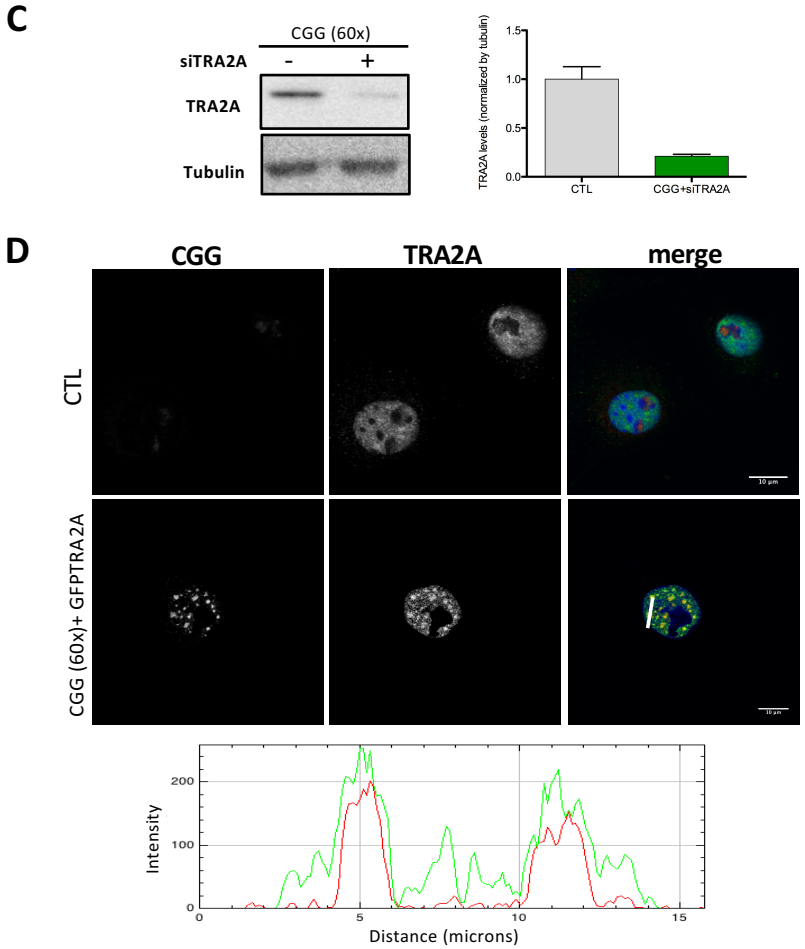
Zou, W., Chen, X., Shim, J.-H., Huang, Z., Brady, N., Hu, D., Drapp, R., Sigrist, K., Glimcher, L.H., and Jones, D. (2011). The E3 ubiquitin ligase Wwp2 regulates craniofacial development through mono-ubiquitylation of Goosecoid. Nat. Cell Biol. *13*, 59–65.

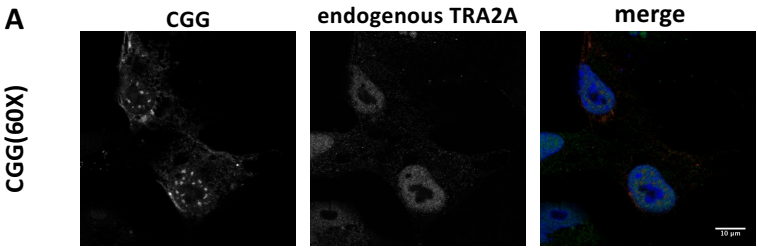

A
Network Property
Degree
Closeness
Betweenness
Prot-Prot
Prot-RNA
Prot-Prot
Prot-RNA
Yeast
Human
Network
P
0.1 0.2 0.3 0.4
B
RNA targets (log scale)
p = 0.003
Granule RBPs
Non granule RBPs
C
Fraction of RBP pairs
Granule forming pairs (Human)
Non–granule RBP pairs (Human)
Granule forming pairs (Yeast)
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

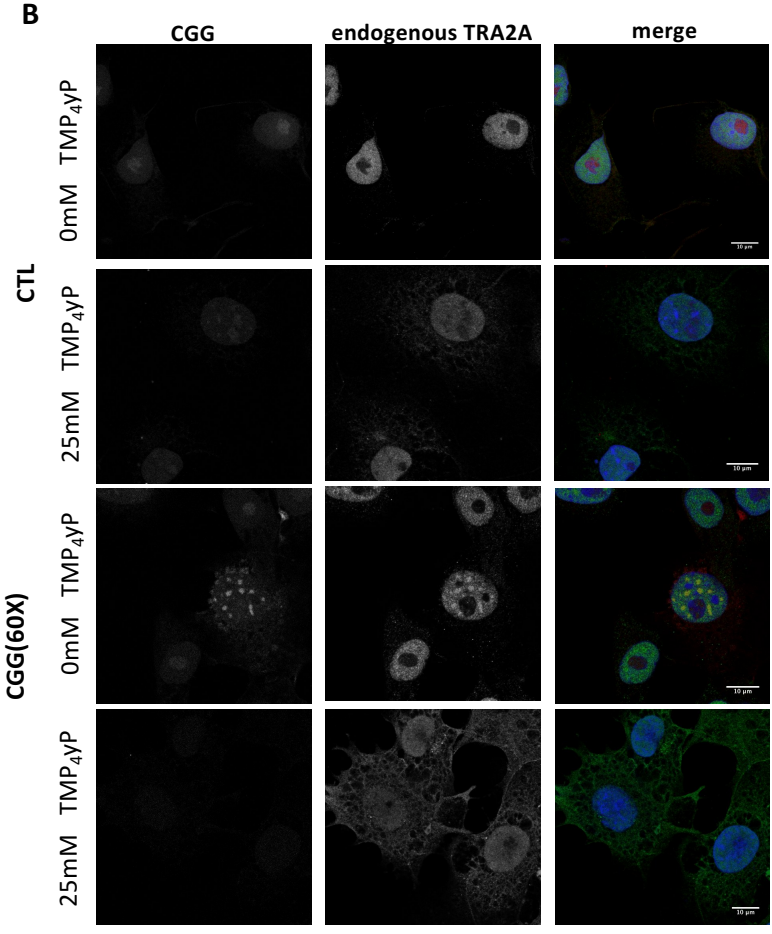

A
Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
−1
0
1
2
p = 5e-11
Transcript Relative Expression (Log 10 RPKM)
Granule
Non−granule
B
Cumulative Distribution Function
1.00
0.75
0.50
0.25
0.00
0.1
0.2
0.3
0.4
0.5
p = 0.005
Structural Content (PARS)
Granule
Non−granule
C
Cumulative Distribution 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

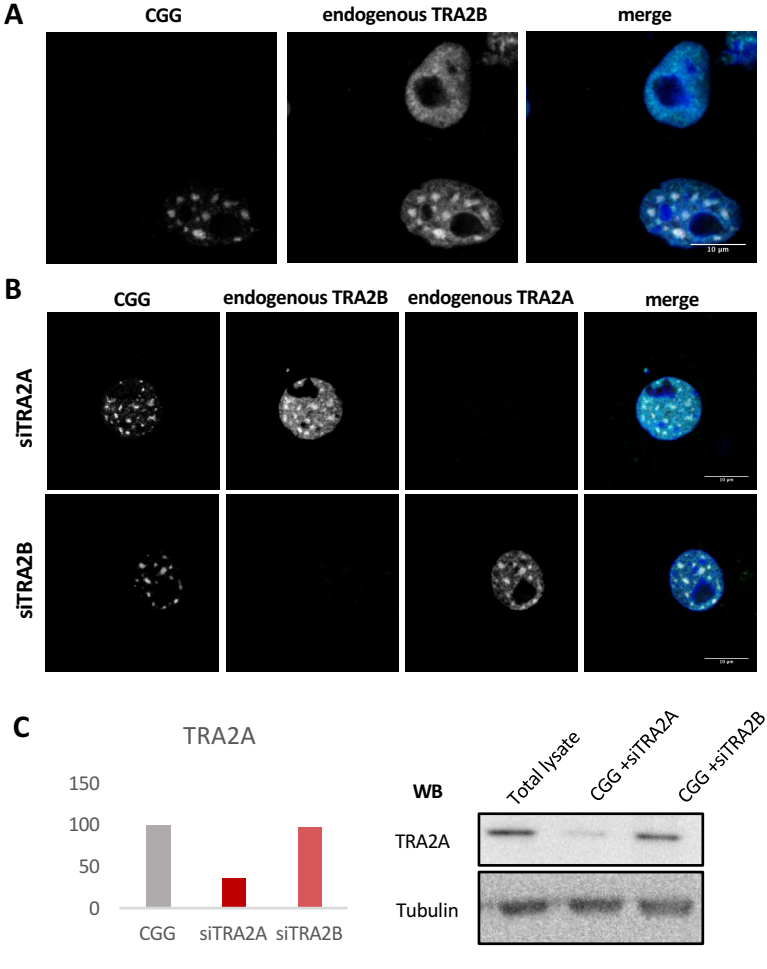

A
TRA2A
Validated in vitro
Interaction Score (DP)
Specificity (IS)
0.00
0.25
0.50
0.75
1.00
B
correlation=0.75
Signal Fold Change (Replica 2)
Signal Fold Change (Replica 1)
0
2
4
6
C
PRE
WT
Predictive Power (AUC)
RBP Ranking by Affinity
0.55
0.60
0.65
0.70
0.75
0.80
15
25
35
45
D
Granule Propensity
0.00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

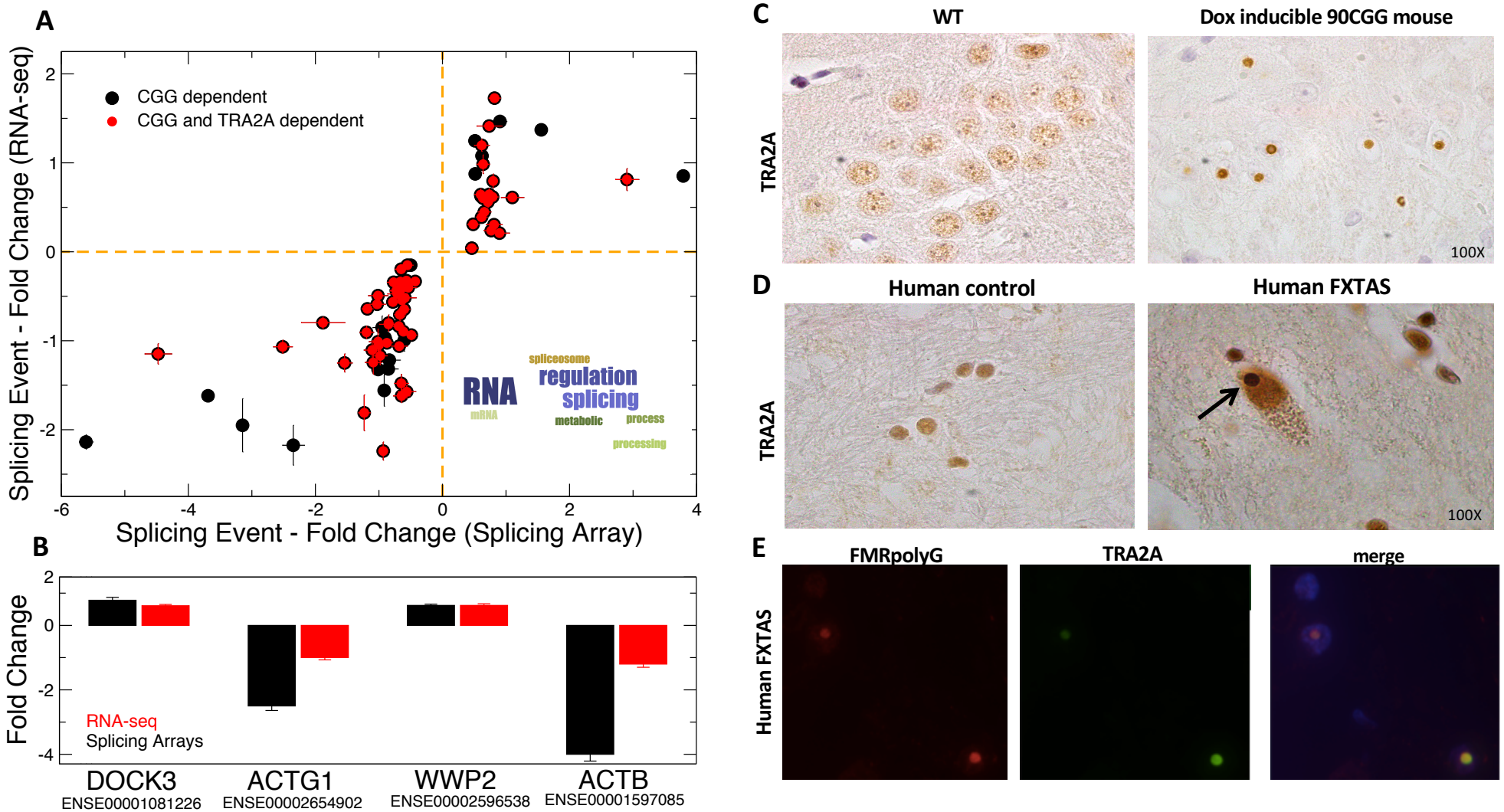

A
CGG
endogenous TRA2A
merge
CTL
CGG (60)x
Intensity
Distance (microns)
B
CGG
TRA2A
merge
CTL
siTRA2A
10 µm


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
CGG(60X)
10 µm


B
CGG
endogenous TRA2A
merge
CTL
CGG(60X)
0mM TMP4yP
25mM TMP4yP
0mM TMP4yP
25mM TMP4yP
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)
RNA
mRNA
spliceosome
regulation
splicing
metabolic
process
processing
B
Fold Change
RNA-seq
Splicing Arrays
DOCK3
ENSE00001081226
ACTG1
ENSE00002654902
WWP2
ENSE00002596538
ACTB
ENSE00001597085
C
WT
Dox inducible 90CGG mouse
TRA2A
100X
D
Human control
Human FXTAS
TRA2A
100X
E
FMRpolyG
TRA2A
merge
Human FXTAS